

GEODESY FOR THE LAYMAN
DEFENSE MAPPING AGENCY
BUILDING 56 U S NAVAL OBSERVATORY
DMA TR 80-003
WASHINGTON DC 20305

FOR THE DIRECTOR:
VIRGIL J JOHNSON
Captain, USN
Chief of Staff
Converted to L^AT_EX
by Jonny L. Gent 2016

16 March 1984

Forward

- I. The basic principles of geodesy are presented in an elementary form. The formation of geodetic datums is introduced and the necessity of connecting or joining datums is discussed. Methods used to connect independent geodetic systems to a single world reference system are discussed, including the role of gravity data. The 1983 edition of this publication contains an expanded discussion of satellite and related technological applications to geodesy and an updated description of the World Geodetic System.
- II. The Defense Mapping Agency is not responsible for publishing revisions or identifying the obsolescence of its technical publications.
- III. DMA TR 80-003 contains no copyrighted material, nor is a copyright pending. This publication is approved for public release; distribution unlimited. Reproduction in whole or in part is authorized for U.S. Government use. Copies may be requested from the Defense Technical Information Center, Cameron Station Alexandria, VA 22314.

Defense Mapping Agency

The Defense Mapping Agency provides mapping, charting and geodetic support to the Secretary of Defense, the Joint Chiefs of Staff, the military departments and other Department of Defense components. The support includes production and worldwide distribution of maps, charts, precise positioning data and digital data for strategic and tactical military operations and weapon systems. The Defense Mapping Agency also provides nautical charts and marine navigational data for the worldwide merchant marine and private yachtsmen.

The agency is under the direction and control of the Under Secretary of Defense for Research and Engineering and its director is responsible to the Joint Chiefs of Staff for operational matters.

The Defense Mapping Agency maintains close liaison with civilian agencies of the U.S. government which are engaged in mapping, charting and geodetic (MC&G) activities. To further the advances in MC&G, the agency works with various national and international scientific and operational organizations.

Preface

GEODESY FOR THE LAYMAN, first published in 1959, was written by Lt Col Richard K. Burkard (retired). Since that time and through this, its fourth revision, a host of authors have revised and supplemented the original work. The publication attempts to develop an appreciation of the scope of geodesy and to generate some understanding of the problems encountered in the science. The basic principles are presented through discussions of various geodetic topics such as: ellipsoids, geoid, horizontal surveying, leveling, geodetic datums and gravity. This edition provides a revised chapter on Satellite Geodesy to present the results of growth in this field and a new chapter on Other Developments in Geodesy to present the effect of recent applications of new technology to

Geodesy. Also, the chapter on the World Geodetic System has been extensively revised and placed last in keeping with the World Geodetic System's role as the end product of much of the endeavor described in the preceding chapters.

Contents

1	History of Geodesy	5
2	Figure of the Earth	8
3	Geodetic Surveying Techniques	12
4	Geodetic Systems	23
5	Physical Geodesy	36
6	Satellite Geodesy	44
7	Other Developments in Geodesy	54
8	The World Geodetic System	59
9	Bibliography	64
A	Spherical Harmonics	67

Introduction

What is geodesy? Who needs it and why? These are some of the questions asked by many people. Actually, geodesy is nothing new having been around for centuries. Webster defines geodesy as THAT BRANCH OF APPLIED MATHEMATICS WHICH DETERMINES BY OBSERVATION AND MEASUREMENT THE EXACT POSITIONS OF POINTS AND THE FIGURES AND AREAS OF LARGE PORTIONS OF THE EARTH'S SURFACE, THE SHAPE AND SIZE OF THE EARTH, AND THE VARIATIONS OF TERRESTRIAL GRAVITY. It is a specialized application of several familiar facets of basic mathematical and physical concepts. In practice, geodesy uses the principles of mathematics, astronomy and physics, and applies them within the capabilities of modern engineering and technology. A thorough study of the science of geodesy is not a simple undertaking. However, it is possible to gain an understanding of the historical development, a general knowledge of the methods and techniques of the science, and the way geodesy is being used to solve some Department of Defense (DoD) problems.

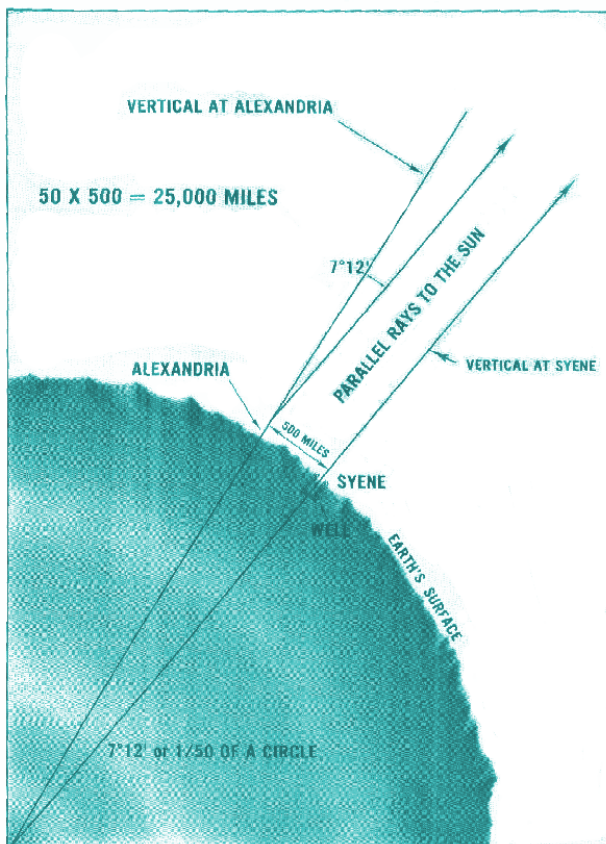
In the past, military geodesy was largely involved with the practical aspect of the determination of exact positions of points on the earth's surface for mapping or artillery control purposes while the determination of the precise size and shape of the earth was a purely scientific role. However, modern requirements for distance and direction require both the practical and scientific applications of the science to provide the answers to problems in such fields as satellite tracking, global navigation and defensive missile operations.

Chapter 1

History of Geodesy

Man has been concerned about the earth on which he lives for many centuries. During very early times this concern was limited, naturally, to the immediate vicinity of his home; later it expanded to the distance of markets or exchange places; and finally, with the development of means of transportation man became interested in his whole world. Much of this early WORLD INTEREST was evidenced by speculation concerning the size, shape, and composition of the earth.

Figure 1



The early Greeks, in their speculation and theorizing, ranged from the flat disc advocated by Homer to Pythagoras' spherical figure - an idea supported one hundred years later by Aristotle. Pythagoras was a mathematician and to him the most perfect figure was a sphere. He reasoned that the gods would create a perfect figure and therefore the earth was created to be spherical in shape. Anaximenes, an early Greek scientist, believed strongly that the earth was rectangular in shape.

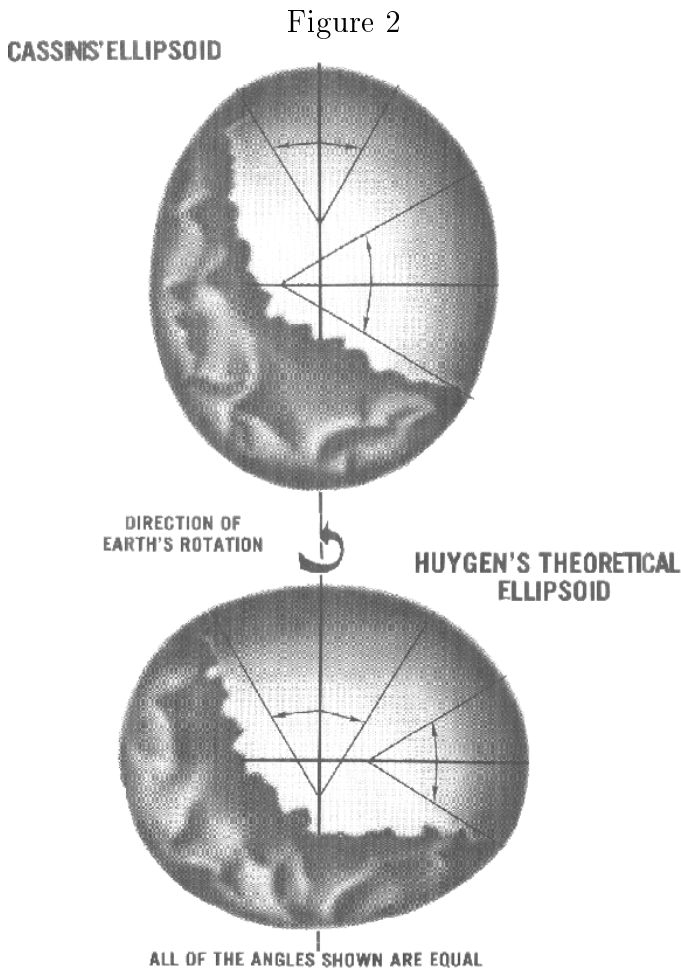
Since the spherical shape was the most widely supported during the Greek Era, efforts to determine its size followed. Plato determined the circumference of the earth to be 40,000 miles while Archimedes estimated 30,000 miles. Plato's figure was a guess and Archimedes' a more conservative approximation. Meanwhile, in Egypt, a Greek scholar and philosopher, Eratosthenes, set out to make more explicit measurements.

He had observed that on the day of the summer solstice, the midday sun shone to

the bottom of a well in the town of Syene (Aswan).

At the same time, he observed the sun was not directly overhead at Alexandria; instead, it cast a shadow with the vertical equal to $\frac{1}{50}^{th}$ of a circle ($7^{\circ} 12'$). To these observations, Eratosthenes applied certain KNOWN facts (1) that on the day of the summer solstice, the midday sun was directly over the line of the summer Tropic Zone (Tropic of Cancer)-Syene was therefore concluded to be on this line; (2) the linear distance between Alexandria and Syene was 500 miles; (3) Alexandria and Syene lay on a direct north south line.

From these observations and KNOWN facts, Eratosthenes concluded that, since the angular deviation of the sun from the vertical at Alexandria was also the angle of the subtended arc, the linear distance between Alexandria and Syene was $\frac{1}{50}^{th}$ of the circumference of the earth or $50 \times 500 = 25,000$ miles. A currently accepted value for the earth's circumference at the Equator is 24,901 miles, based upon the equatorial radius of the World Geodetic System (Chapter 8). The actual unit of measure used by Eratosthenes was called the 'stadia'. No one knows for sure what the stadia that he used is in today's units. The measurements given above in miles were derived using one stadia equal to one-tenth statute mile.



It is remarkable that such accuracy was obtained in view of the fact that most of the KNOWN facts and his observations were incorrect:

- I. although it is true that the sun at noon is directly overhead at the Tropic of Cancer on the day of the summer solstice, it was erroneously concluded that Syene lay on the line.
- II. Actually, Syene is 37 miles to the north;
- III. the true distance between Alexandria and Syene is 453 miles and not 500;
- IV. Syene lies $3^{\circ} 30'$ east of the meridian of Alexandria;
- V. the difference of latitude between Alexandria and Syene is $7^{\circ} 5'$ rather than $7^{\circ} 12'$ as Eratosthenes had concluded.

Another ancient measurement of the size of the earth was made by the Greek, Posidonius. He noted that a certain star was hidden from view in most parts of Greece but that it just grazed the horizon at Rhodes. Posidonius measured the elevation of the same star at Alexandria and determined that the angle was $1/48^{th}$ of

circle. Assuming the distance from Alexandria to Rhodes to be 500 miles, he computed the circumference of the earth as 24,000 miles. While both his measurements were approximations when combined, one error compensated for another and he achieved a fairly accurate result.

Revising the figures of Posidonius, another Greek philosopher determined 18,000 miles as the earth's circumference. This last figure was promulgated by Ptolemy through his world maps. The maps of Ptolemy strongly influenced the cartographers of the middle ages. It is probable that Columbus, using such maps, was led to believe that Asia was only 3 or 4 thousand miles west of Europe. It was not until the 15th century that his concept of the earth's size was revised. During that period the Flemish cartographer, Mercator, made successive reductions in the size of the Mediterranean Sea and all of Europe which had the effect of increasing the size of the earth.

The telescope, logarithmic tables, and the method of triangulation were contributed to the science of geodesy during the 17th century. In the course of the century, the Frenchman, Picard, performed an arc measurement that is modern in some respects. He measured a base line by the aid of wooden rods, used a telescope in his angle measurements, and computed with logarithms. Cassini later continued Picard's arc northward to Dunkirk and southward to the Spanish boundary. Cassini divided the measured arc into two parts, one northward from Paris, another southward. When he computed the length of a degree from both chains, he found that the length of one degree in the northern part of the chain was shorter than that in the southern part. This unexpected result could have been caused only by an egg-shaped earth or by observational errors.

The results started an intense controversy between French and English scientists. The English claimed that the earth must be flattened, as Newton and Huygens had shown theoretically, while the Frenchmen defended their own measurement and were inclined to keep the earth egg-shaped.

To settle the controversy, once and for all, the French Academy of Sciences sent a geodetic expedition to Peru in 1735 to measure the length of a meridian degree close to the Equator and another to Lapland to make a similar measurement near the Arctic Circle. The measurements conclusively proved the earth to be flattened, as Newton had forecast. Since all the computations involved in a geodetic survey are accomplished in terms of a mathematical surface (reference ellipsoid) resembling the shape of the earth, the findings were very important.

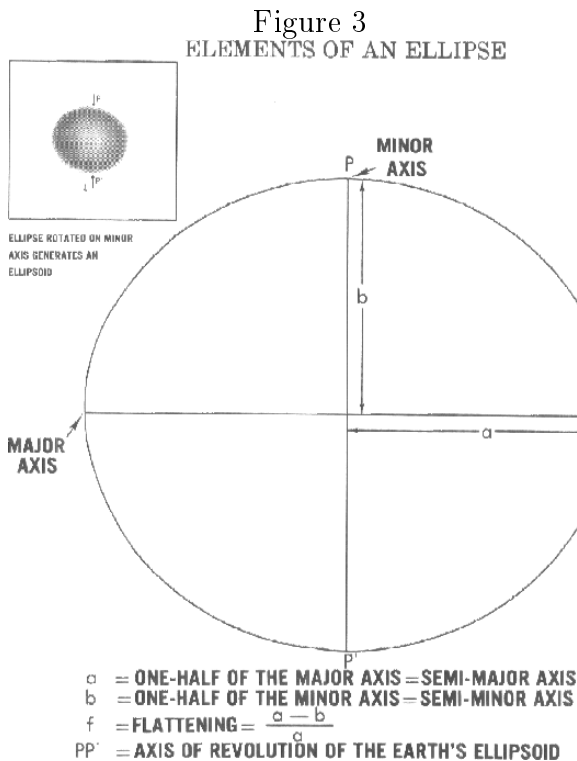
Chapter 2

Figure of the Earth

The expression FIGURE OF THE EARTH has various meanings in geodesy according to the way it is used and the precision with which the earth's size and shape is to be defined. The actual topographic surface is most apparent with its variety of land forms and water areas. This is, in fact, the surface on which actual earth measurements are made. It is not suitable, however, for exact mathematical computations because the formulas which would be required to take the irregularities into account would necessitate a prohibitive amount of computations. The topographic surface is generally the concern of topographers and hydrographers.

The Pythagorean spherical concept offers a simple surface which is mathematically easy to deal with. Many astronomical and navigational computations use it as a surface representing the earth.

While the sphere is a close approximation of the true figure of the earth and satisfactory for many purposes, to the geodesists interested in the measurement of long distances-spanning continents and oceans-a more exact figure is necessary. The idea of flat earth, however, is still acceptable for surveys of small areas. Plane-table surveys are made for relatively small areas and no account is taken of the curvature of the earth. A survey of a city would likely be computed as though the earth were a plane surface the size of the city. For such small areas, exact positions can be determined relative to each other without considering the size and shape of the total earth.



Ellipsoid of Revolution

Since the earth is in fact flattened slightly at the poles and bulges somewhat at the equator, the geometrical figure used in geodesy to most nearly approximate the shape of the earth is an ellipsoid of revolution. The ellipsoid of revolution is the figure which would be obtained by rotating

an ellipse about its shorter axis.

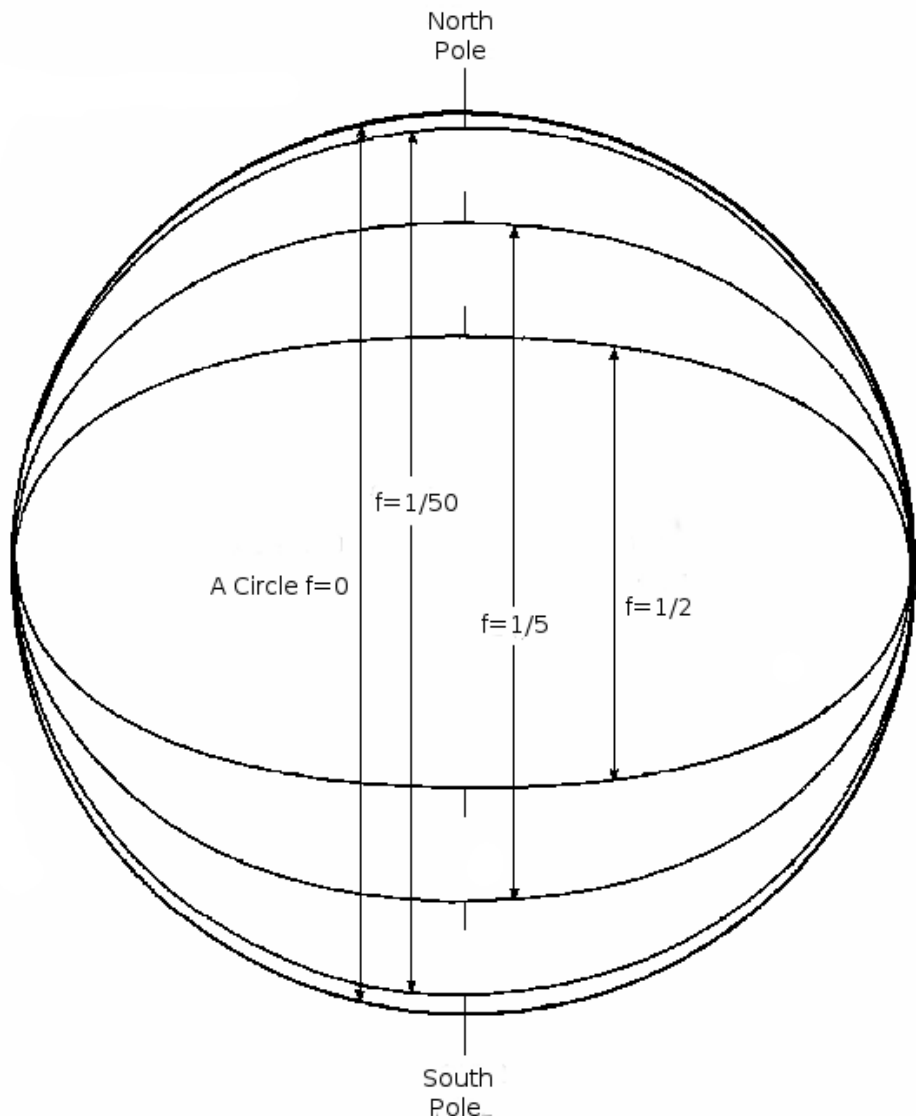
An ellipsoid of revolution is uniquely defined by specifying two dimensions. Geodesists, by convention, use the semi-major axis and flattening. The size is represented by the radius at the equator, the semi-major axis and designated by the letter, 'a.' The shape of the ellipsoid is given by the flattening, f , which indicates how closely an ellipsoid approaches a spherical shape. The difference between the ellipsoid of revolution representing the earth and a sphere is very small.

The ellipsoids listed below have had utility in geodetic work and many are still in use. The older ellipsoids are named for the individual who derived them and the year of development is given. The international ellipsoid was developed by Hayford in 1910 and adopted by the International Union of Geodesy and Geophysics (IUGG) which recommended it for international use.

At the 1967 meeting of the IUGG held in Lucerne, Switzerland, the ellipsoid called GRS-67 in the listing was recommended for adoption. The new ellipsoid was not recommended to replace the International Ellipsoid (1924), but was advocated for use where a greater degree of accuracy is required. It became a part of the Geodetic Reference System 1967 which was approved and adopted at the 1971 meeting of the IUGG held in Moscow. It is used in Australia for the Australian Geodetic Datum and in South America for the South American Datum 1969.

Figure 4

The Earth's Flattening is about 1/300



The ellipsoid called GRS80 (Geodetic Reference System 1980) was approved and adopted at the 1979 meeting of the IUGG held in Canberra, Australia. The ellipsoids used to define WGS66 and

WGS72 are discussed in Chapter 8.

NAME	YEAR	EQUATORIAL RADIUS m	FLATTENING	WHERE USED
Krassowsky	(1940)	6,378,245	1/298.30	Russia
International	(1924)	6,378,388	1/297.00	Europe
Clarke	(1880)	6,378,249	1/293.46	France, Africa
Clarke	(1866)	6,378,206	1/294.98	North America
Bessel	(1841)	6,377,397	1/299.15	Japan
Airy	(1830)	6,377,563	1/299.32	Great Britain
Everest	(1830)	6,377,276	1/300.80	India
WGS66	(1966)	6,378,145	1/298.25	USA/DoD
GRS67	(1967)	6,378,160	1/298.25	Australia South America
WGS72	(1972)	6,378,135	1/298.26	USA/DoD
GRS80	(1979)	6,378,137	1/298.26	NAD83

The possibility that the earth's equator is an ellipse rather than a circle and therefore that the ellipsoid is triaxial has been a matter of scientific controversy for many years. Modern technological developments have furnished new and rapid methods for data collection and since the launching of the first Russian sputnik, orbital data has been used to investigate the theory of ellipticity.

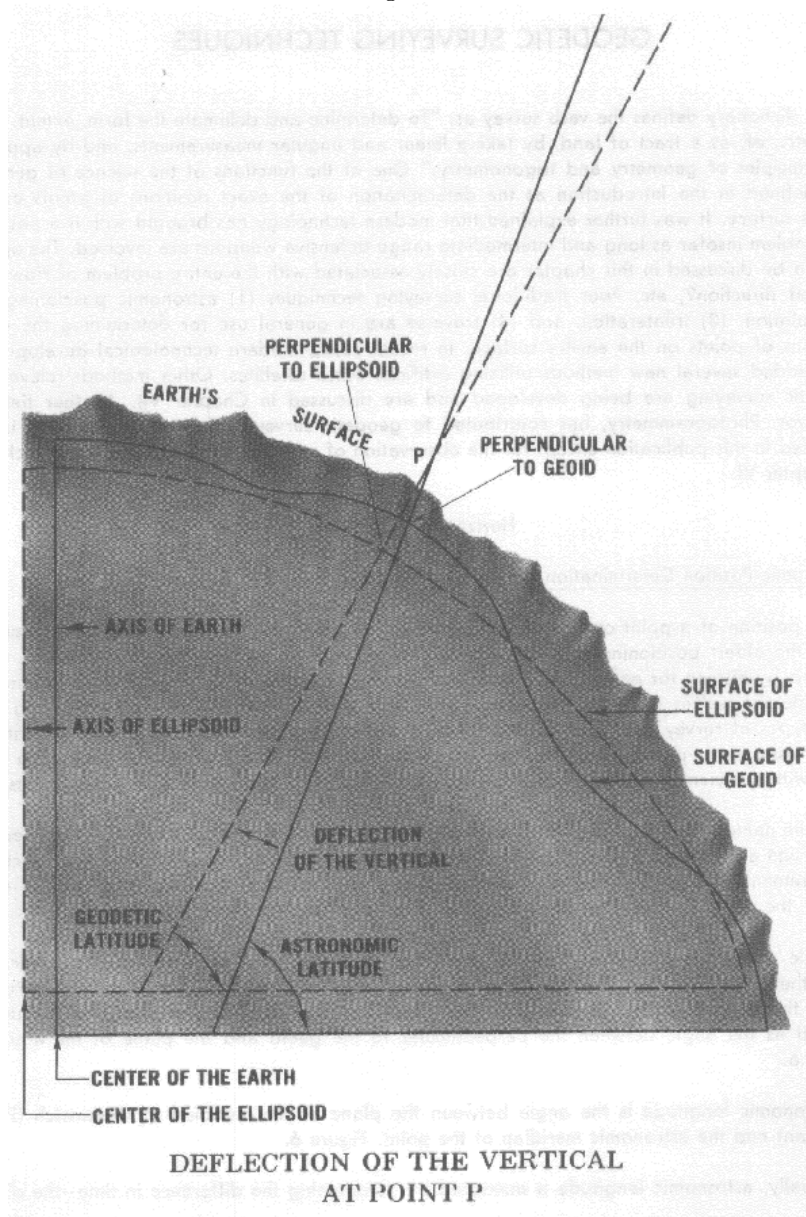
A second theory, more complicated than triaxiality, proposed that satellite orbital variations indicate additional flattening at the south pole accompanied by a bulge of the same magnitude at the north pole. It is also contended that the northern middle latitudes were slightly flattened and the southern middle latitudes bulged in a similar amount. This concept suggested a slight pear shaped earth and was the subject of much public discussion. Modern geodesy tends to retain the ellipsoid of revolution and treat triaxiality and pear shape as a part of the geoid separation (to be discussed later).

Geoid

It was stated earlier that measurements are made on the apparent or topographic surface of the earth and it has just been explained that computations are performed on an ellipsoid. One other surface is involved in geodetic measurement - the geoid. In geodetic surveying, the computation of the geodetic coordinates of points is performed on an ellipsoid which closely approximates the size and shape of the earth in the area of the survey. The actual measurements made on the surface of the earth with certain instruments are referred to the geoid, as explained below. The ellipsoid is a mathematically defined regular surface with specific dimensions. The geoid, on the other hand, coincides with that surface to which the oceans would conform over the entire earth if free to adjust to the combined effect of the earth's mass attraction and the centrifugal force of the earth's rotation. As a result of the uneven distribution of the earth's mass, the geoidal surface is irregular and, since the ellipsoid is a regular surface, the two will not coincide. The separations are referred to as geoid undulations, geoid heights, or geoid separations.

The geoid is a surface along which the gravity potential is everywhere equal and to which the direction of gravity is always perpendicular. The latter is particularly significant because optical instruments containing leveling devices are commonly used to make geodetic measurements. When properly adjusted, the vertical axis of the instrument coincides with the direction of gravity and is, therefore, perpendicular to the geoid. The angle between the plumb line which is perpendicular to the geoid (sometimes called THE VERTICAL) and the perpendicular to the ellipsoid (sometimes called THE NORMAL) is defined as the deflection of the vertical. Figure 5 shows the north-south component of the deflection of the vertical.

Figure 5



Chapter 3

Geodetic Surveying Techniques

The dictionary defines the verb SURVEY as, TO DETERMINE AND DELINEATE THE FORM, EXTENT, POSITION, ETC., OF, AS A TRACT OF LAND, BY TAKING LINEAR AND ANGULAR MEASUREMENTS, AND BY APPLYING THE PRINCIPLES OF GEOMETRY AND TRIGONOMETRY. One of the functions of the science of geodesy was defined in the Introduction as the determination of the exact positions of points on the earth's surface. It was further explained that modern technology has brought with it a positioning problem insofar as long and intermediate range defensive weapons are involved. The operations to be discussed in this chapter are closely associated with the entire problem of HOW FAR? IN WHAT DIRECTION?, ETC. Four traditional surveying techniques:

- I. astronomic positioning;
- II. triangulation;
- III. trilateration; and
- IV. traverse.

are in general use for determining the exact positions of points on the earth's surface. In recent years, modern technological developments have added several new methods utilizing artificial earth satellites. Other methods relevant to geodetic surveying are being developed and are discussed in Chapter 7. Another field of endeavor, Photogrammetry, has contributed to geodetic surveying for many years but is not discussed in this publication except for the observation of satellites by cameras which is included in Chapter 6.

Horizontal Positioning

Astronomic Position Determination

The position of a point can be obtained directly by observing the stars. Astronomic positioning is the oldest positioning method. It has been used for many years by mariners and, more recently, by airmen for navigational purposes. Explorers have often used the astronomic method to locate themselves in uncharted areas. Geodesists must use astronomic positions along with other types of survey data such as triangulation and trilateration to establish precise positions. Single astronomic

positions not interconnected by geodetic surveys cannot be related to each other with sufficient accuracy for the computation of distance and direction between points.

As the name implies, astronomic positions are obtained by measuring the angles between the plumb line at the point and a star or series of stars and recording the precise time at which the measurements are made. After combining the data with information obtained from star catalogs, the direction of the plumb line (zenith direction) is computed.

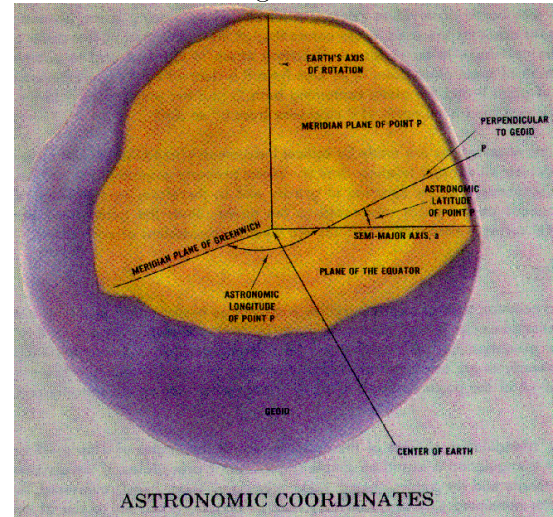
While geodesists use elaborate and very precise techniques for determining astronomic latitude, the simplest method, in the northern hemisphere, is to measure the elevation of Polaris above the horizon of the observer. For the purposes of this publication, astronomic latitude is defined as the angle between the perpendicular to the geoid and the plane of the equator.

Astronomic longitude is the angle between the plane of the Prime Meridian at Greenwich and the astronomic meridian of the point. Actually, astronomic longitude is measured by determining the difference in time—the difference in hours, minutes, and seconds between the time a specific star is directly over the Greenwich meridian and the time the same star is directly over the meridian plane of the point. Shortwave radio equipment is used to obtain time signals which can be referred to Greenwich Mean Time while chronometers (very accurate clocks) are used to measure the time at the point. By referring to a star catalog, the exact Greenwich Mean Time the star was over the Prime Meridian is obtained. The difference between the time at the point and the time at Greenwich is used to compute the astronomic longitude of the point. Since a point of the earth rotates through 360° in 24 hours, the difference in local time between two points can be easily converted into difference in longitude.

Another astronomic observation related to horizontal positioning is the astronomic azimuth. Very accurate azimuths are used in the controlling of the orientation of first-order triangulation which is the next topic to be discussed. Referring again to Figure 6 and to point P, the astronomic azimuth of some other point Q as seen from P is defined as the angle between the meridian plane of point P and the plane containing both Q and the perpendicular to the geoid at P. This angle is reckoned from north at P clockwise from O to 360° .

Astronomic observations are made by optical instruments such as a theodolite, zenith camera, prismatic astrolabe, all containing leveling devices. When properly adjusted, the instrument's vertical axis coincides with the direction of gravity and is, therefore, perpendicular to the geoid. Thus, astronomic positions are referenced to the geoid. Since the geoid is an irregular, non-mathematical surface, astronomic positions are wholly independent of each other.

Figure 6



Triangulation

The most common type of geodetic survey is known as triangulation. It differs from the plane survey mentioned

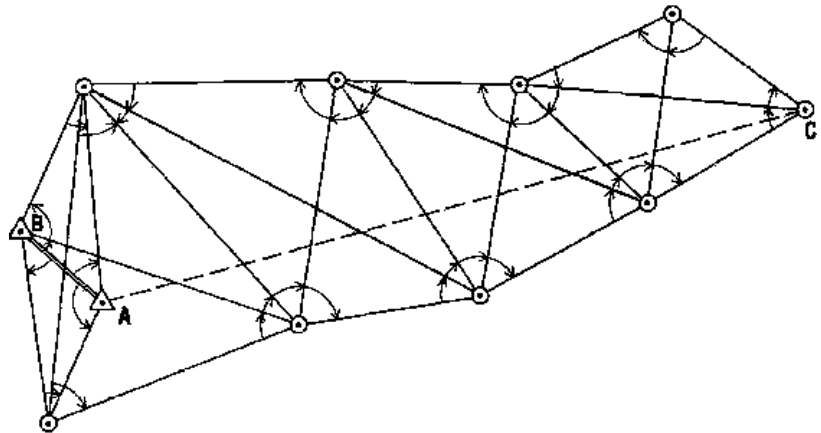
earlier in that more accurate instruments are used, instrumental errors are either removed or predetermined so that they can be compensated for in the computations and more rigorous procedures are employed to reduce observational errors. Another very important difference is that all of the positions established by triangulation are mathematically related to each other.

Basically, triangulation consists of the measurement of the angles of a series of triangles. The principle of triangulation is based on simple trigonometric procedures. If the distance along one side of a triangle and the angles at each end of the side are accurately measured, the other two sides and the remaining angle can be computed. Normally, all of the angles of every triangle are measured for the minimization of error and to furnish data for use in computing the precision of the measurements. Also, the latitude and longitude of one end of the measured side along with the length and direction (azimuth) of the side provide sufficient data to compute the latitude and longitude of the other end of the side.

The measured side of the base triangle is called a base line. Measurements are made as carefully and accurately as possible with specially calibrated tapes or wires of invar, an alloy highly resistant to changes in length resulting from changes in temperature. The tapes or wires are checked periodically against standard measures of length (at the Bureau of Standards in the United States and corresponding agencies in other countries). The geodimeter and tellurometer, operating on electro-optical and electronic principles respectively, have replaced the older methods of base measurement in the recent surveys. The work can be completed more rapidly and accurately than with wire or tape. The laser equipped geodimeter has proven to be the most accurate and it can measure much longer distances without losing accuracy.

To establish an arc of triangulation between two widely separated locations, a base line may be measured and longitude and latitude determined for the initial point at one end. The locations are then connected by a series of adjoining triangles forming quadrilaterals extending from each end.

Figure 7



Known Data:
Length of baseline AB.
Latitude and longitude of Points A and B.
Azimuth of line AB.

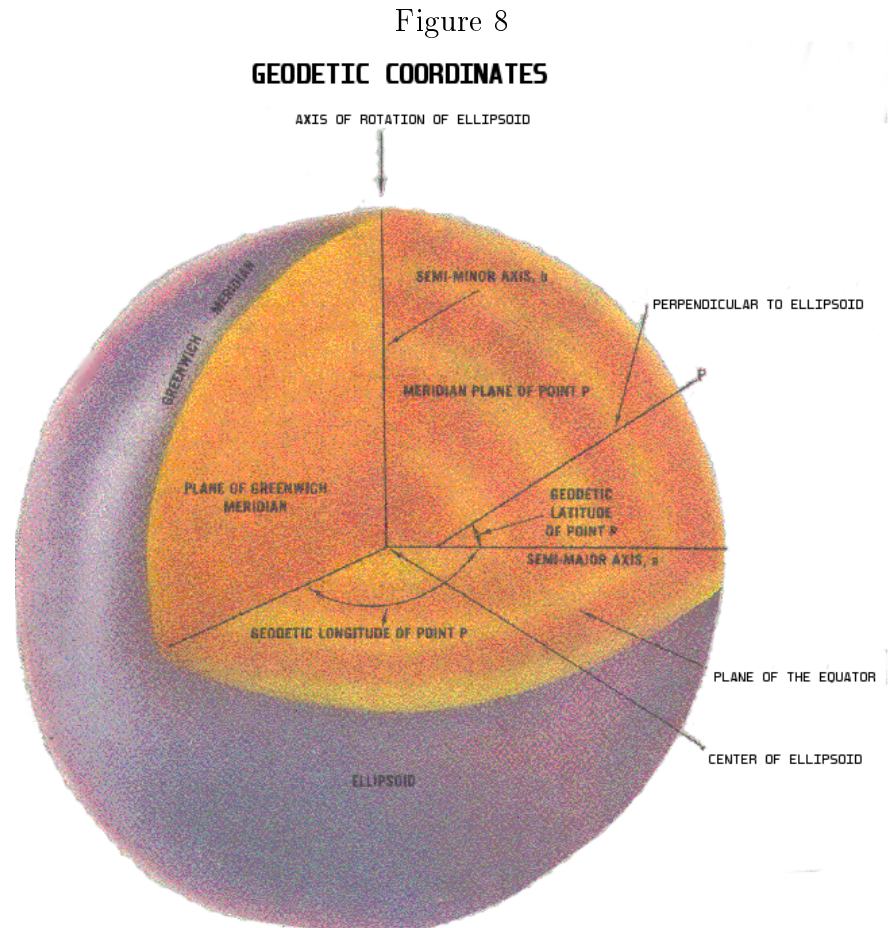
Measured Data:
Angles to new control points.

Computed Data:
Latitude and Longitude of point C and other new points.
Length and azimuth of line AC.
Length and azimuth of all other lines.

A Simple Triangulation Net

With the longitude, latitude, and azimuth of the initial points, similar data is computed for each vertex of the triangles thereby establishing triangulation stations or geodetic control stations. The coordinates of each of the stations are defined as geodetic coordinates.

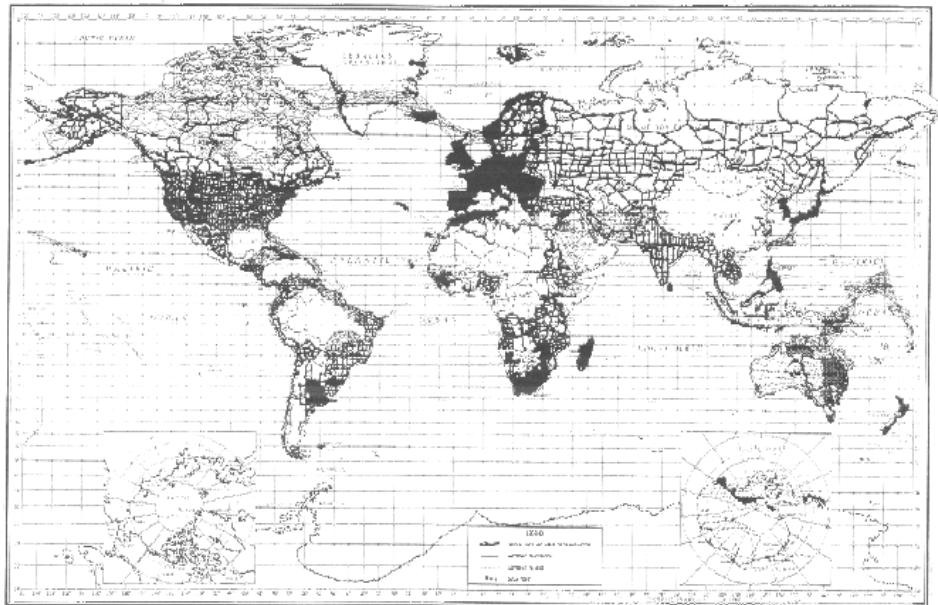
Triangulation is extended over large areas by connecting and extending series of arcs and forming a network or triangulation system. The network is adjusted in a manner which reduces the effect of observational errors to a minimum. A denser distribution of geodetic control is achieved in a system by subdividing or filling in with other surveys. Figure 9 serves to illustrate, in a general manner, the major triangulation networks which have been established.



There are four general orders of triangulation.

First-Order (Primary Horizontal Control) is the most accurate triangulation. It is costly and time-consuming using the best instruments and rigorous computation methods. First-Order triangulation is usually used to provide the basic framework of horizontal control for a large area such as for a national network. It has also been used in preparation for metropolitan expansion and for scientific studies requiring exact geodetic data. Its accuracy should be at least one part in 100,000.

Figure 9



Second-Order, Class I (Secondary Horizontal Control) includes the area networks between the First-Order arcs and detailed surveys in very high value land areas. Surveys of this class strengthen the US National Horizontal Control Network and are adjusted as part of the network. Therefore, this class also includes the basic framework for further densification. The internal closures of Second-Order, Class I triangulation should indicate an accuracy of at least one part in 50,000. The demands for reliable horizontal control surveys in areas which are not in a high state of development or where no such development is anticipated in the near future justifies the need for a triangulation classified as Second-Order, Class II (Supplemental Horizontal Control). This class is used to establish control along the coastline, inland waterways and interstate highways. The control data contributes to the National Network and is published as part of the network. The minimum accuracy allowable in Class II of Second-Order is one part in 20,000.

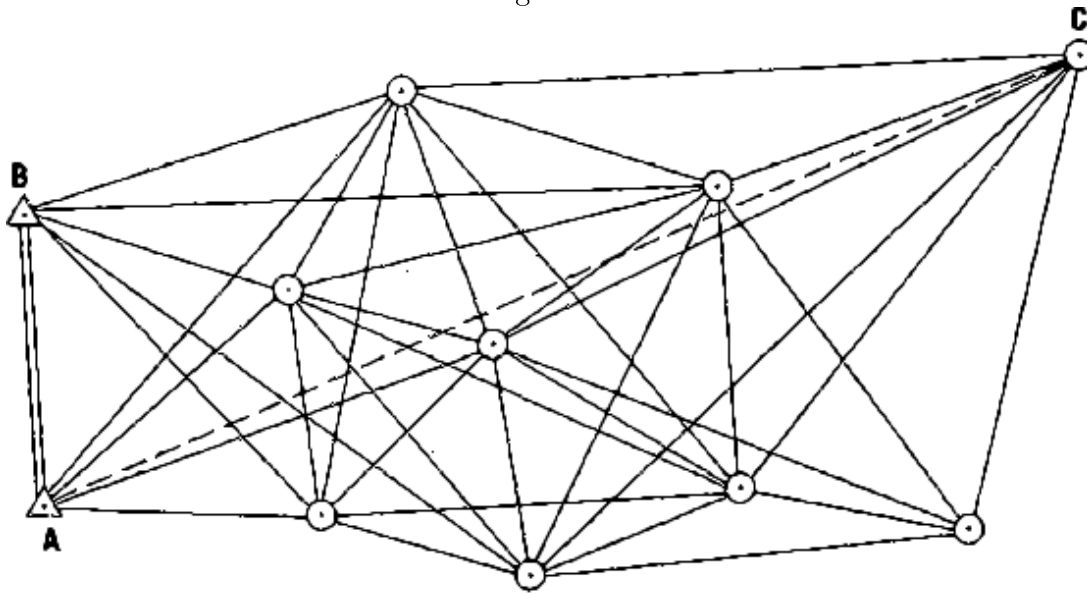
Third-Order, Class I and Class II (Local Horizontal Control) is used to establish control for local improvements and developments, topographic and hydrographic surveys, or for such other projects for which they provide sufficient accuracy. This triangulation is carefully connected to the National Network. The work should be performed with sufficient accuracy to satisfy the standards of one part in 10,000 for Class I and one part in 5,000 for Class II. Spires, stacks, standpipes, flag poles and other identifiable objects located to this accuracy also have significant value for many surveying and engineering projects.

The sole accuracy requirement for Fourth-Order triangulation is that the positions be located without any appreciable errors on maps compiled on the basis of the control.

Normally, triangulation is carried out by parties of surveyors occupying preplanned locations (stations) along the arc and accomplishing all the measurements as they proceed. When distances between two points were too long for conventional methods, connections were sometimes made by a method known as flare triangulation. Stations were occupied on either side of the gap and

magnesium flares were parachuted from aircraft or 'shot' into the air from ships at suitable points between them. Intersections of lines were made simultaneously at all of the stations and reasonably accurate BRIDGES established. A connection of this type was established between Norway and Denmark. However, satellite geodesy (Chapter 6) has solved the problem of bridging wide gaps.

Figure 10



KNOWN DATA:

- Length of baseline AB
- Latitude and longitude of points A and B
- Azimuth of line AB

MEASURED DATA:

- Length of all triangle sides

COMPUTED DATA:

- Latitude and longitude of C and all other points
- Length and azimuth of line AC
- Length and azimuth between any two points

A TRILATERATION NET

Trilateration

Another surveying method that has been used involves the use of radar and aircraft. The SHORAN, HIRAN and SHIRAN electronic distance measuring systems have been applied to performing geodetic surveys by a technique known as trilateration.

Since very long lines (to 500 miles) could be measured by these systems, geodetic triangulation networks have been extended over vast areas in comparatively short periods of time. In addition, the surveys of islands and even continents separated by extensive water barriers have been connected by the techniques. The Canadian SHORAN network connecting the sparsely populated northern coastal and island areas with the central part of the country and the North Atlantic HIRAN Network tying North America to Europe are examples of the application of the trilateration technique.

Figure 11 shows these and several other trilateration networks (SHORAN and HIRAN) which have been established throughout the world. SHIRAN has been used in the interior of Brazil.

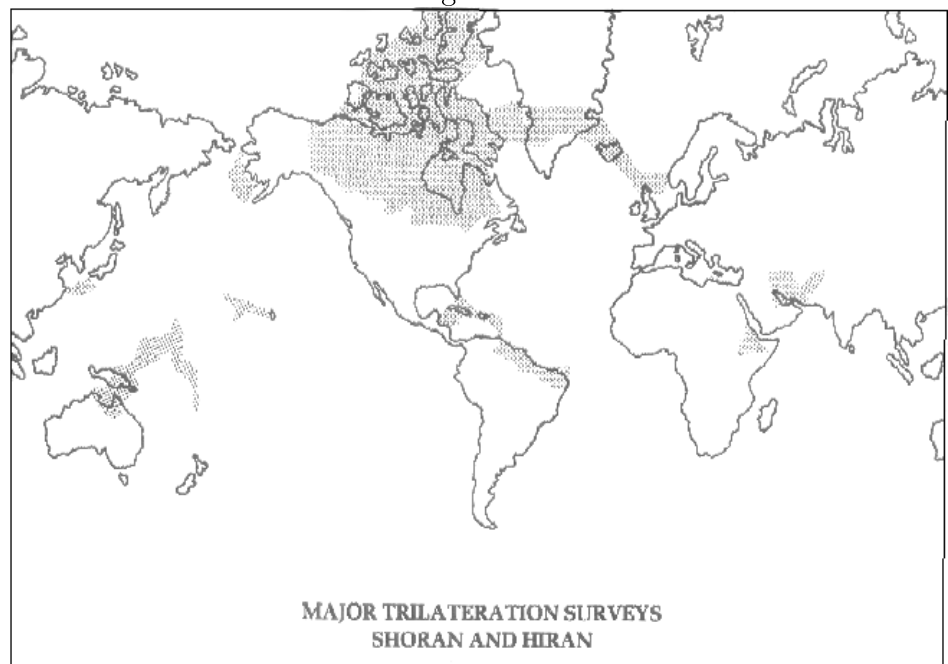
Only distances are measured in trilateration and each side is measured repeatedly to insure precision. The entire network is then adjusted to minimize the effects of the observations errors. The angles of the triangles are computed so the geodetic positions are obtained as in triangulation.

Traverse

The simplest method of extending control is called traverse. The system is similar to dead reckoning navigation where distances and directions are measured. In performing a traverse,

the surveyor starts at a known position with a known azimuth (direction) to another point and measures angles and distances between a series of survey points. With the angular measurements, the direction of each line of the traverse can be computed; and with the measurements of the length of the lines, the position of each control point computed. If the traverse returns to the starting point or some other known position, it is a closed traverse, otherwise the traverse is said to be open.

Figure 11



Since electronic distance measuring equipment has become available, the accuracy of traverse surveys has increased significantly. The tellurometer (microwave) has been used in Australia to complete a network (Australian Geodetic Datum) covering that continent. The average loop length

is about 900 miles; the average loop closure of this work is 2.2 parts per million. The laser equipped geodimeter has been used to produce internal accuracies better than one part per million in establishing the transcontinental traverse in the United States. The traverse consists of a series of high-precision length, angle and astronomic azimuth determinations running approximately east-west and north-south through the conterminous states, forming somewhat rectangular loops. Figure 13. This traverse will be the BACKBONE of a re-adjustment of the horizontal control network in this country.

Celestial Techniques

Celestial methods in geodesy involves the determination of an observer's position from observations of the moon, stars and satellites. Celestial triangulation permits the extension of long arcs across oceans and inaccessible space terrain. Satellites also permit a determination of the shape of the earth and provide important knowledge of its gravitational field (discussion of satellite geodesy is reserved for Chapter 6). All of the celestial methods possess one common characteristic-observed data is not affected by the direction of the vertical at the observation point.

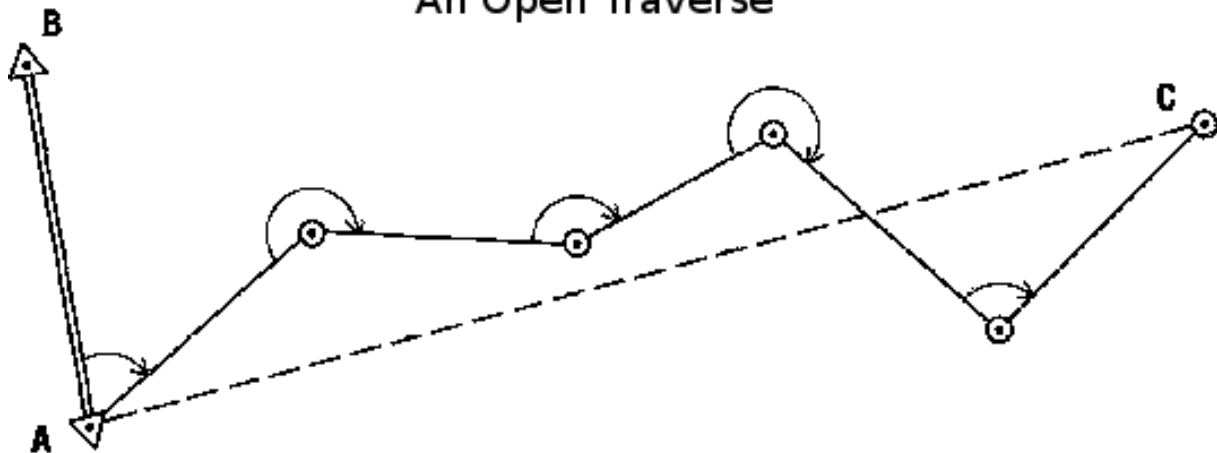
Geodetic experiments have been performed in the past with the solar eclipse, star occultation and moon-position camera methods, but for various reasons the experiments did not produce useful geodetic results. The three techniques have been replaced by the observation and tracking of artificial earth satellites.

Vertical Positioning

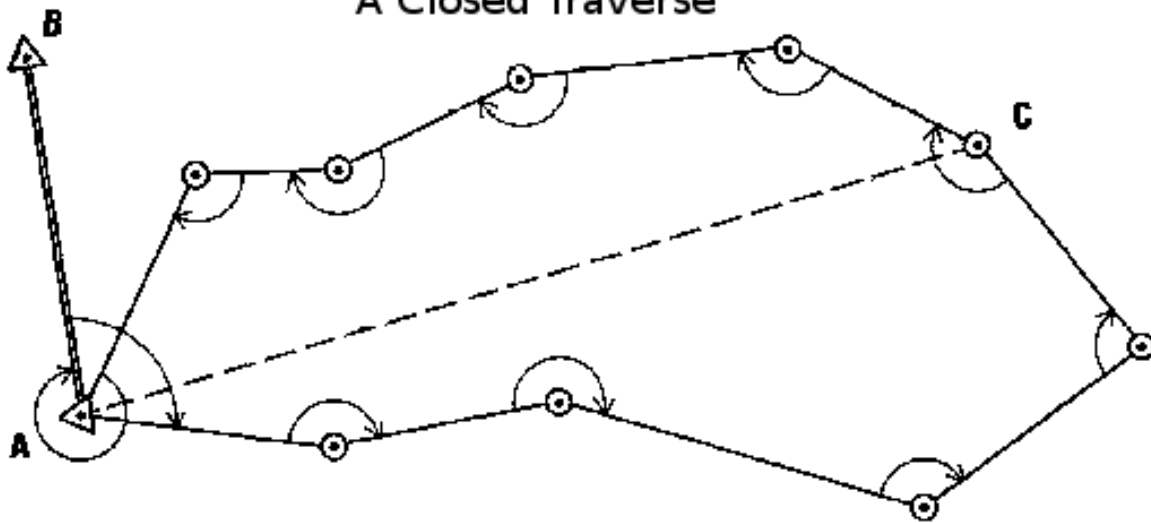
Vertical surveying is the process of determining heights-elevations above the mean sea level surface. As noted earlier, the geoid corresponds to the mean level of the open sea. In geodetic surveys executed primarily for mapping purposes, there is no problem in the fact that geodetic positions are referred to an ellipsoid and the elevations of the positions are referred to the geoid. However, geodetic data for missiles requires an adjustment in the elevation information to compensate for the undulations of

the geoid above and below the regular mathematical surface of the ellipsoid. The adjustment uses complex advanced geodetic techniques. One method based on Stokes' Theorem is mentioned in the discussion of physical geodesy (Chapter 5).

Figure 12
An Open Traverse



A Closed Traverse



Known Data:

- Latitude and longitude of Point A.
- Azimuth of line AB.

Measured Data:

- Length of traverse sides.
- Angles between traverse sides

Computed Data

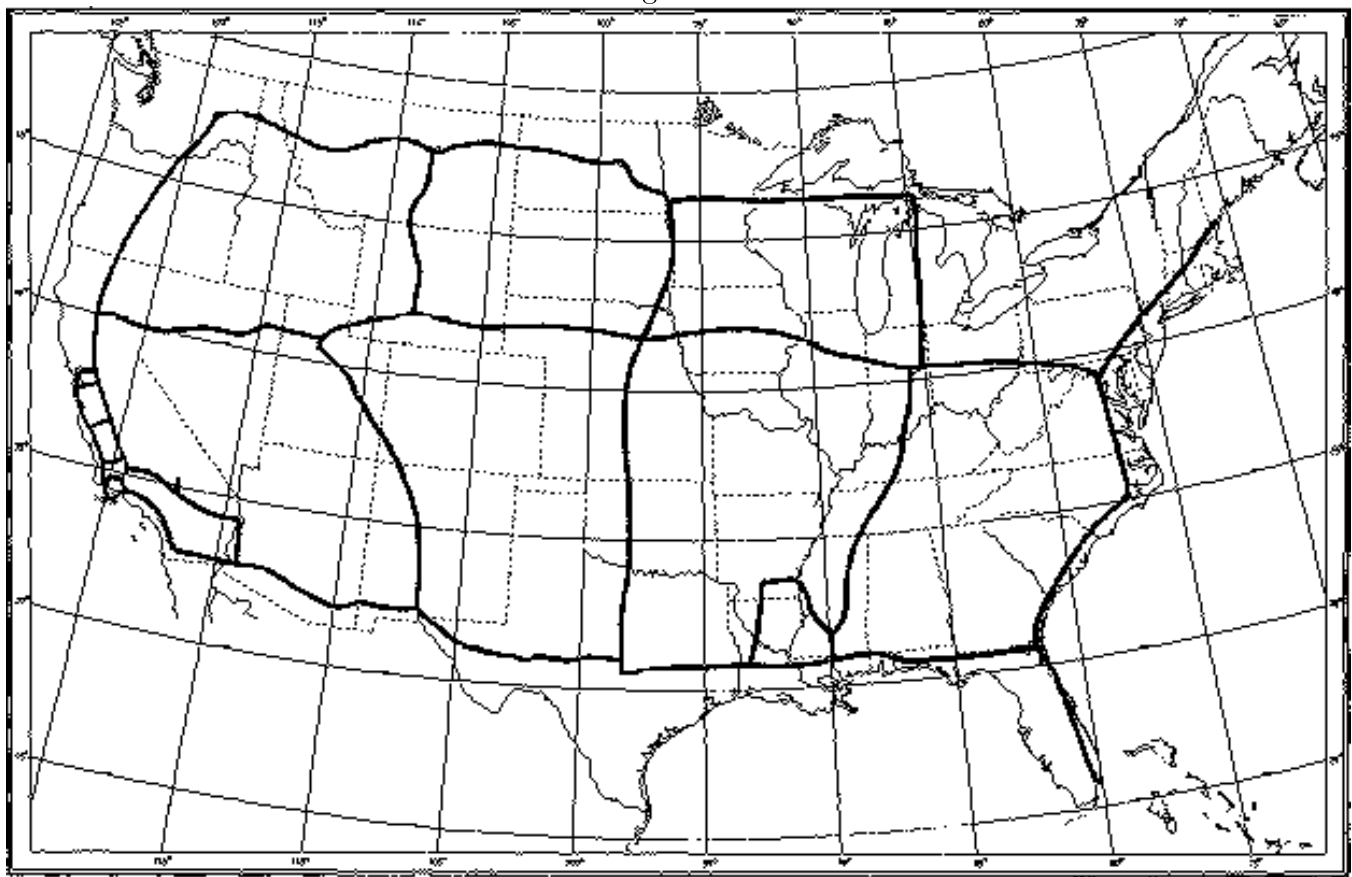
- Latitude and longitude of point C and other points.
- Length and azimuth of line AC.
- Length and azimuth of line between any two points.

Precise geodetic leveling is used to establish a basic network of vertical control points. From these, the height of other positions in the survey can be determined by supplementary methods. The mean sea level surface used as a reference (vertical datum) is determined by obtaining an average of the hourly water heights for a period of several years at tidal gauges.

There are three leveling techniques-differential, trigonometric, and barometric-which yield information of varying accuracy.

Differential leveling is the most accurate of the three methods. With the instrument locked in position, readings are made on two calibrated staffs held in an upright position ahead of and behind the instrument. The difference between readings is the difference in elevation between the points.

Figure 13

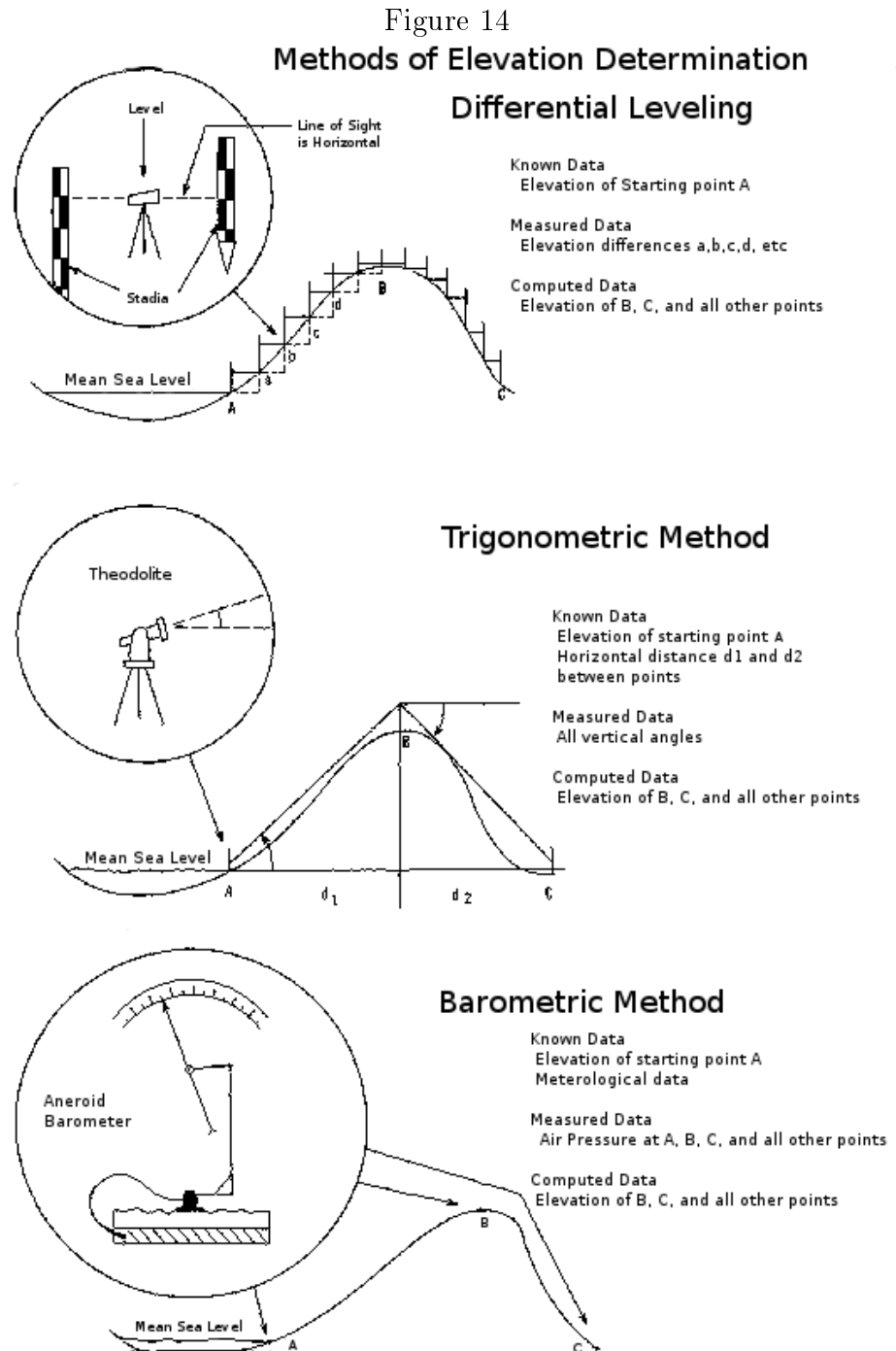


Transcontinental Traverse

The optical instrument used for leveling contains a bubble tube to adjust it in a position parallel to the geoid. When properly 'set-up' at a point, the telescope is locked in a perfectly horizontal (level) position so that it will rotate through a 360° arc. The exact elevation of at least one point in a leveling line must be known and the rest computed from it.

Trigonometric leveling involves measuring a vertical angle from a known distance with a theodolite and computing the elevation of the point. With this method, vertical measurements can be made at the same time horizontal angles are measured for triangulation. It is, therefore, a somewhat more economical method but less accurate than differential leveling. It is often the only practical method of establishing accurate elevation control in mountainous areas.

In barometric leveling, differences in height are determined by measuring the difference in atmospheric pressure at various elevations. Air pressure is measured by mercurial or aneroid barometers, or a boiling point thermometer. Although the degree of accuracy possible with this method is not as great as either of the other two, it is a method which obtains relative heights very rapidly at points which are fairly far apart. It is widely used in the reconnaissance and exploratory surveys where more exacting measurements will be made later or are not required.



Chapter 4

Geodetic Systems

A datum is defined as any numerical or geometrical quantity or set of such quantities which serve as a reference or base for other quantities. In geodesy two types of datums must be considered: a horizontal datum which forms the basis for the computations of horizontal control

surveys in which the curvature of the earth is considered, and a vertical datum to which elevations are referred. In other words, the coordinates for points in specific geodetic surveys and triangulation networks are computed from certain initial quantities (datums).

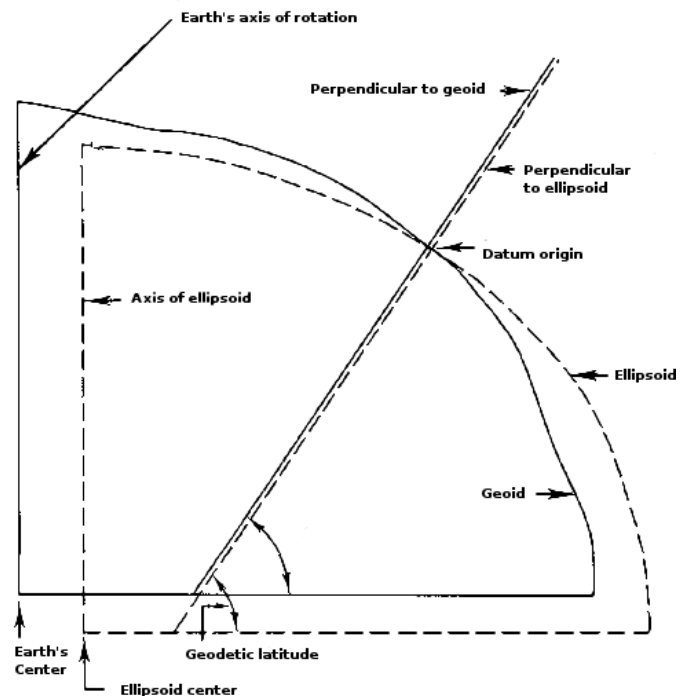
Horizontal Geodetic Datums

A horizontal geodetic datum may consist of the longitude and latitude of an initial point (origin); an azimuth of a line (direction) to some other triangulation station; the parameters (radius and flattening) of the ellipsoid selected for the computations; and the geoid separation at the origin. A change in any of these quantities affects every point on the datum. For this reason, while positions within a system are directly and accurately relatable, data such as distance and azimuth derived from computations involving geodetic positions on different datums will be in error in proportion to the difference in the initial quantities.

Selection of the reference ellipsoid provides the radius and flattening factors. The simplest means of obtaining the other factors to establish the geodetic datum is to select a station, preferably one located near the center of a triangulation network, to serve as the datum origin. Then the astronomical coordinates of the station and the astronomical azimuth of a line from the station

Figure 15
**SINGLE ASTRONOMICAL STATION
DATUM ORIENTATION**

**Perpendicular to ellipsoid made coincident
with perpendicular to geoid at datum origin**



**Orientation of ellipsoid center with respect
to Earth's center**

to another control station are observed. The observed astronomical coordinates and azimuth are adopted without any correction as the geodetic coordinates and azimuth of the datum origin on the reference ellipsoid. Further, the geoid and ellipsoid are assumed to coincide at that point. This means that the deflection of the vertical and the separation between the ellipsoid and geoid are defined as zero at the origin. By using this method of orientation, the normal to the ellipsoid is arbitrarily made to coincide with the plumb line at the datum origin.

Although the computed positions will be correct with respect to each other in this type of orientation, the entire net will be shifted with respect to the axis of the earth. This is not significant for local use of the positions but may introduce large systematic errors as the survey is expanded.

It should be noted that although the deflection and undulation are defined as zero at the origin, deflections will occur at other positions within the network. When comparing the geodetic latitude and longitude of any other point in the net with the corresponding astronomical latitude and longitude of that point, differences will appear between the two sets of values.

A datum oriented by a single astronomical point may produce large systematic geoid separations. The ellipsoid is not earth-centered and its rotational axis is not coincident with the axis of the earth. The inconvenience of such an orientation is that the positions derived from different astronomically oriented datums are not directly comparable to each other in any geodetic computation. The Tokyo Datum, discussed later, is an example of this type of datum orientation.

Astronomical-Geodetic Orientation

The deflections of the vertical at a number of Laplace stations can be used for a second type of datum orientation known as the astro-geodetic orientation. A Laplace station is defined as a triangulation or traverse station at which a geodetic (Laplace) azimuth is derived from an astronomic azimuth by use of the Laplace equation. The Laplace equation expresses the relationship between astronomic azimuth, geodetic azimuth and the astronomic longitude and geodetic longitude. Although it is not in the definition, the astronomic latitude is normally observed at each Laplace station. In an orientation of this type, a correction is made at the origin (initial point) which in effect reduces the sum of the squares of the astro-geodetic deflections at all the Laplace stations to a minimum. One of the Laplace stations in the adjustment is arbitrarily selected as the origin.

The deflection of the vertical-the angle between the plumb line and the normal to the ellipsoid-is usually resolved into a north-south component which is equal to the difference between astronomic and geodetic latitude; and an east-west component proportional to the difference between astronomical and geodetic longitude. (Figure 16 shows only the north-south component.) The Laplace equation provides a means of reconciling the azimuth differences resulting from the use of two separate reference surfaces. Laplace equations are introduced into triangulation adjustments to control the azimuth and orient the ellipsoid. Therefore instead of a zero deflection at the origin as with a single astronomic position, there is a deflection of the vertical. Similarly, the geoid separation can be determined at the origin and the ellipsoid re-oriented so that a best average fit is provided for the ellipsoid and the geoid in the area of the Laplace stations used. Consequently, astro-geodetically oriented datums are applicable over larger areas than those oriented by a single astronomic position.

The astronomical-geodetic orientation has the disadvantage that the deflections of the vertical remain relative. If the ellipsoid is changed, the deflections of the vertical will also change. Secondly, it is necessary to assume a specific orientation of the reference ellipsoid with respect to the geoid before computing the astro-geodetic deflections. The orientation is fixed by the initial values of the datum origin from which the geodetic coordinates were computed. Any change in these initial quantities will change the deflection of the vertical at each point. Consequently, the astro-geodetic deflection of the vertical depends upon a specific geodetic datum and the use of geodetic data developed by this method is limited to relatively small areas.

Discrepancies Between Datums

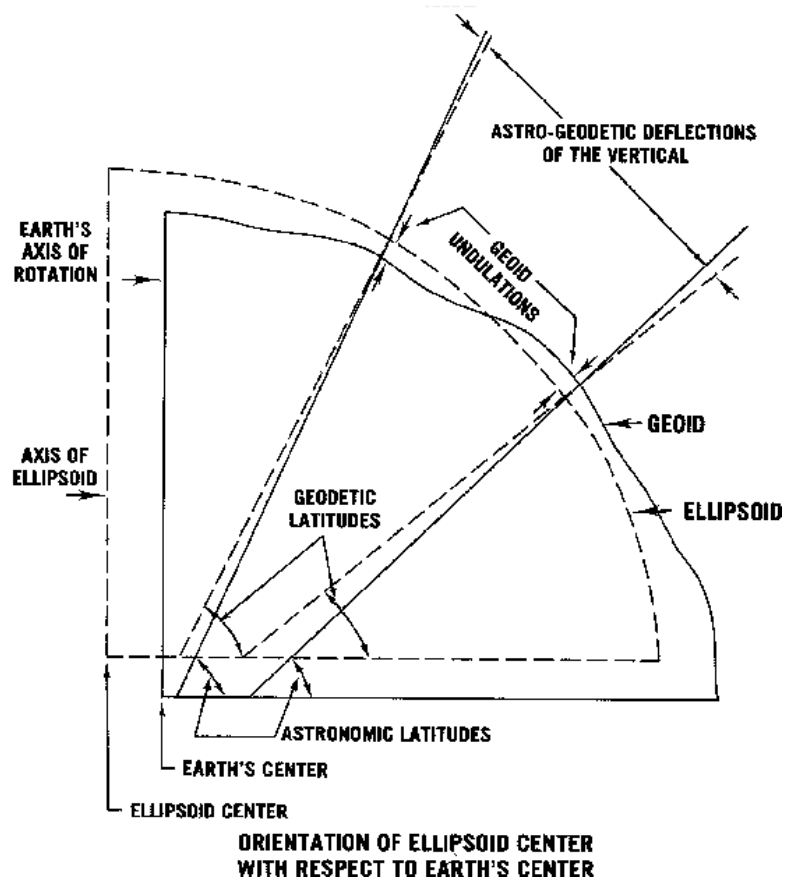
In areas of overlapping geodetic triangulation networks, each computed on a different datum, the coordinates of the points given with respect to one datum will differ from those given with respect to the other. The differences occur because of the different ellipsoids used and the probability that the centers of each datum's ellipsoid is oriented differently with respect to the earth's center. In addition, deflection errors in azimuth cause a relative rotation between the systems. Finally, a difference in the scale of horizontal control may result in a stretch in the corresponding lines of the geodetic nets.

In view of the different orientation of ellipsoid centers, the relative rotation between the systems, and the scale differences; the computation of geodetic information from one datum to another unconnected datum is quite impossible. Regardless of the accuracy of the individual datums for computation within themselves, there is no accurate way to perform distance and azimuth computations between unconnected geodetic systems.

With the development of both intermediate and long range defensive weapon systems, geodetic problems have become more critical than ever before. To satisfy military requirements, it is necessary to provide detailed cartographic coverage of areas of strategic importance and to accomplish geodetic computations between these areas and launch sites which are often on unrelated datums. Both of these requirements necessitate unification of major geodetic datums by one or a combination

Figure 16
Astro-geodetic Datum Orientation

Geoid and ellipsoid are oriented so that the sum of the squares of several deflections of the vertical selected throughout the geodetic network is as small as possible



of existing methods.

Datum Connection

There are three general methods by which horizontal datums can be connected. The first method is restricted to surveys of a limited scope and consists of systematic elimination of discrepancies between adjoining or overlapping triangulation networks. This is done by moving the origin, rotating, and stretching networks to fit each other. The method is usually used to connect local surveys for mapping purposes. This method of datum transformation or datum reduction can only be used where control exists for common points in different systems.

In addition to the classic method of datum transformation described above, there is the gravimetric method of Physical Geodesy (Chapter 5) and the methods of Satellite Geodesy (Chapter 6). These methods are used to relate large geodetic systems to each other or to a world system (Chapter 8). Both the gravimetric and satellite methods produce necessary CONNECTING parameters from reduction of their particular observational data.

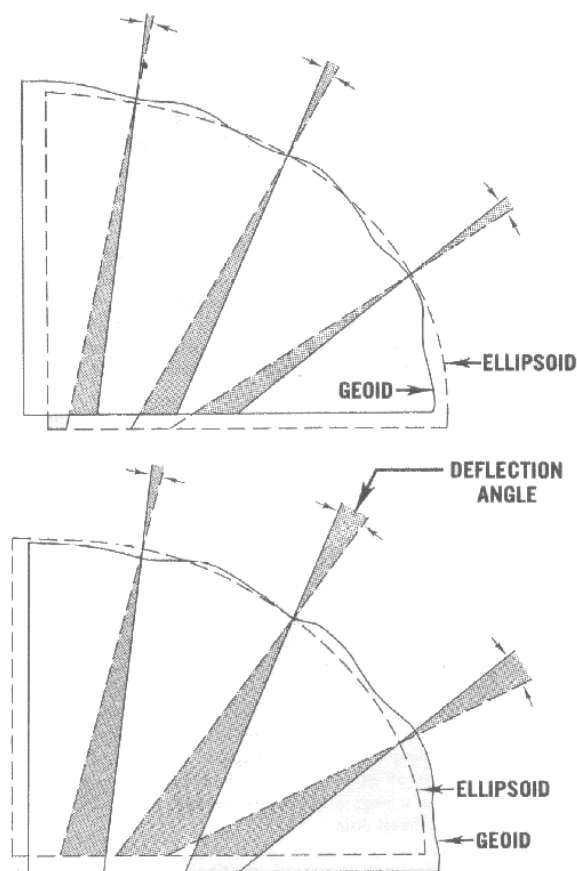
Major Datums Before World War II

By 1940, every technically advanced nation had developed its own geodetic system to an extent governed by its economic and military requirements. Some systems were developed by the expansion and unification of existing local surveys and others by new nationwide surveys replacing outdated local ones. Normally, neighboring countries did not use the same geodetic datum. There was no economic requirement for common geodetic information and the use of

common datums was contrary to the military interests of each country. The only surveys of an international nature based on one datum were the few measurements of long arcs accomplished for the purpose of determining the size and shape of the earth. The net result was that there were many different surveys of varying size which differed from each other remarkably. The national maps based on the surveys also differed widely. Figure 19 illustrates the confusion of geodetic information in Southeast Asia.

As military distance requirements increased, positioning information of local or even national

Figure 17
ASTRO-GEODETIC DEFLECTIONS ARE RELATIVE



NOTE THAT A CHANGE IN ELLIPSOID ORIENTATION
HAS CHANGED THE ASTRO-GEODETIC DEFLECTIONS.

scope became unsatisfactory. The capabilities of the various weapon systems increased until datums of at least continental limits were required.

The Major Datums Today

The best solution (to meeting the increased military distance requirements) was the establishment of a 'single' datum for a large area and adjusting all local systems to it. The North American, European, and Tokyo Datums were (initially) selected for this purpose. (WGS60, Chapter 8). Other major geodetic datums of the world include the Arc and Adindan Datums in Africa, the Australian Geodetic, South American 1969, Indian and the Russian Pulkovo 1942 Datum.

Figure 18

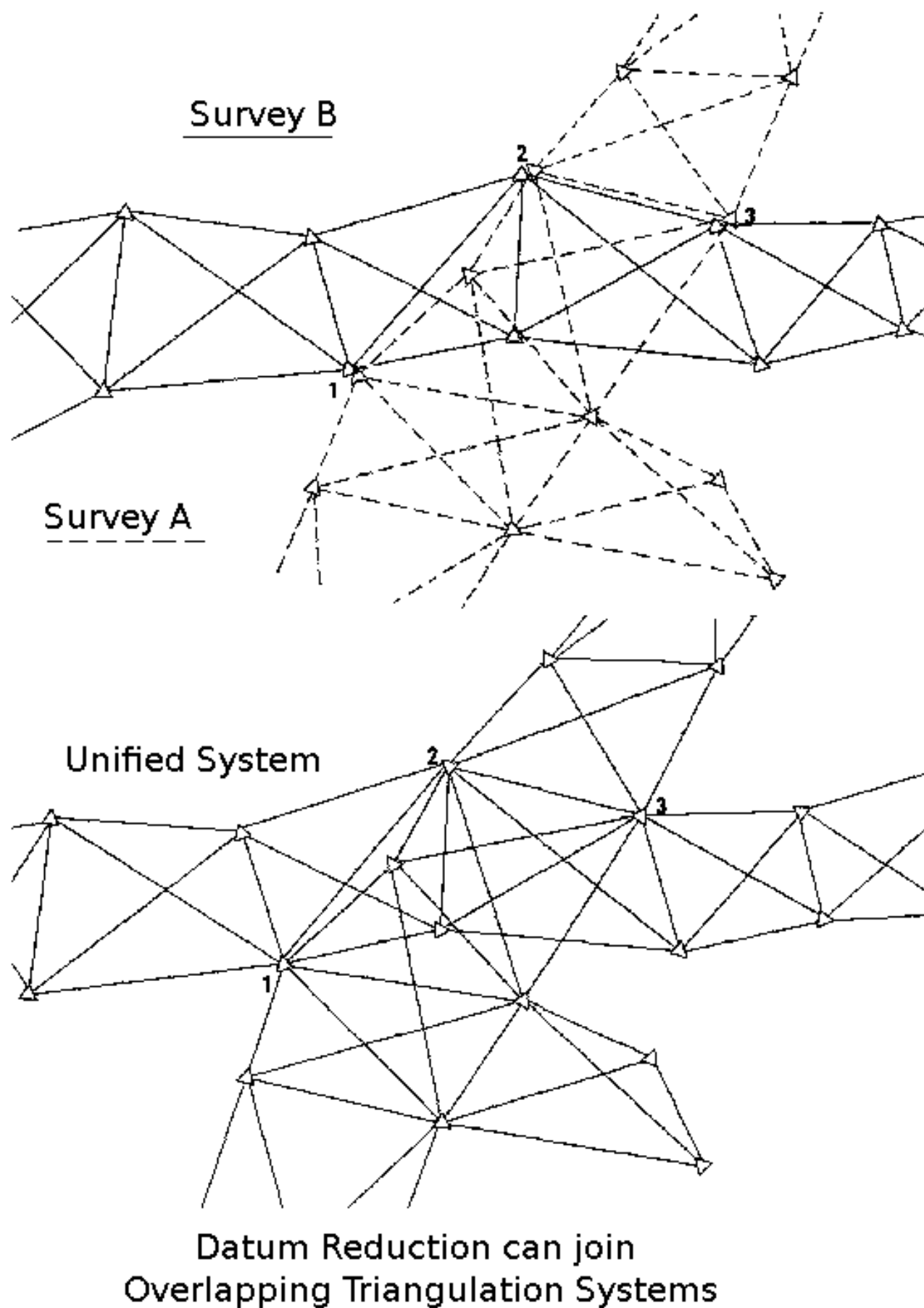


Figure 19

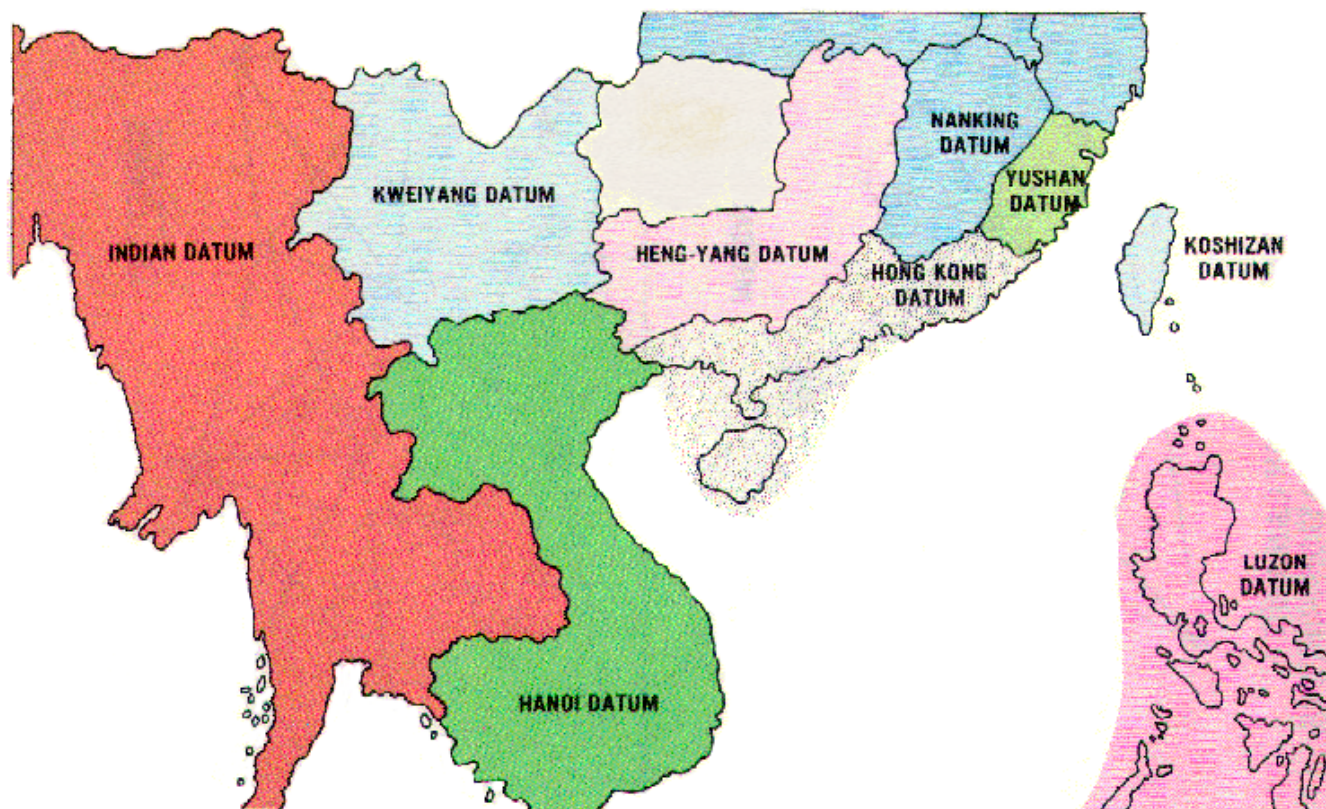


Figure 20

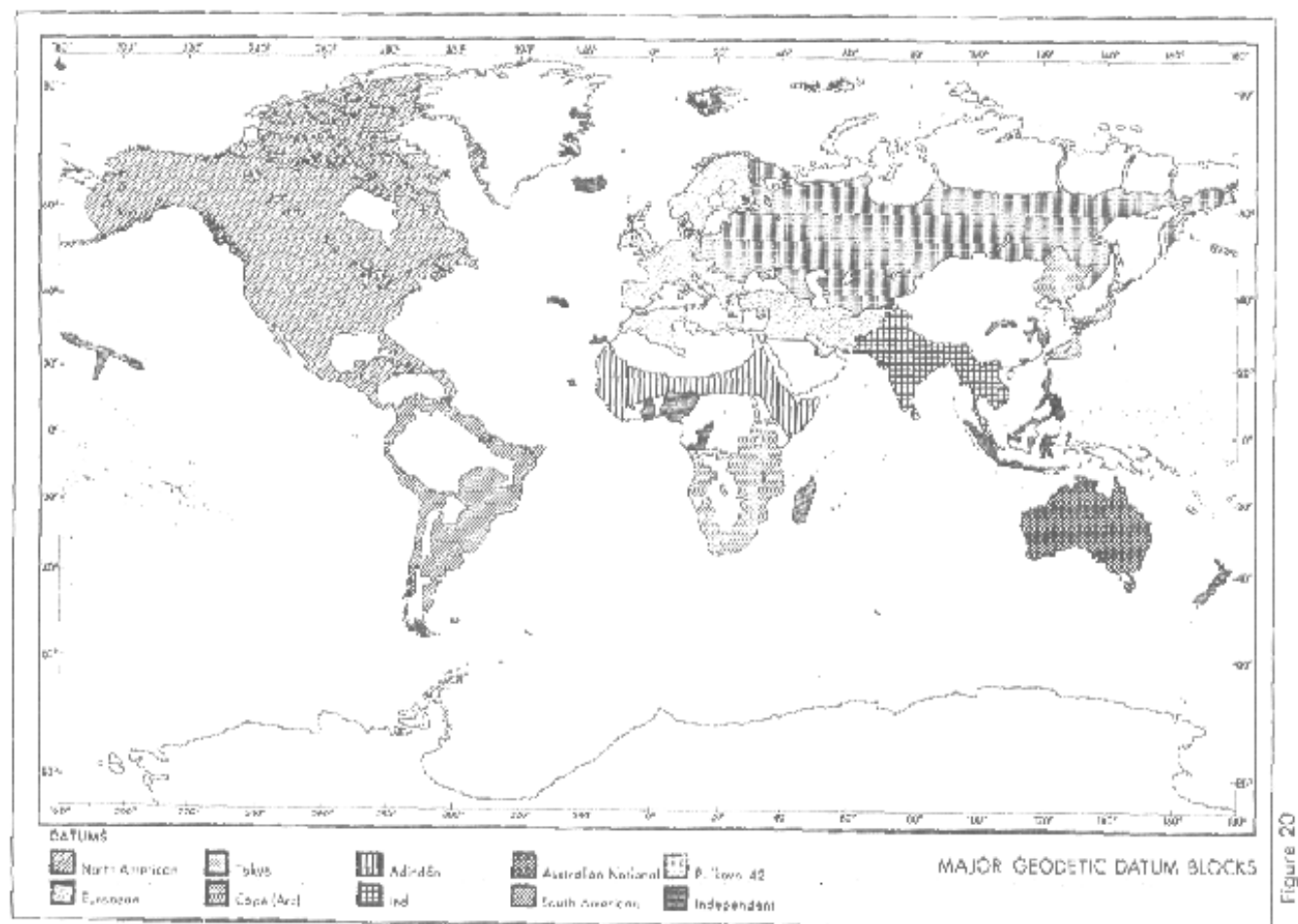
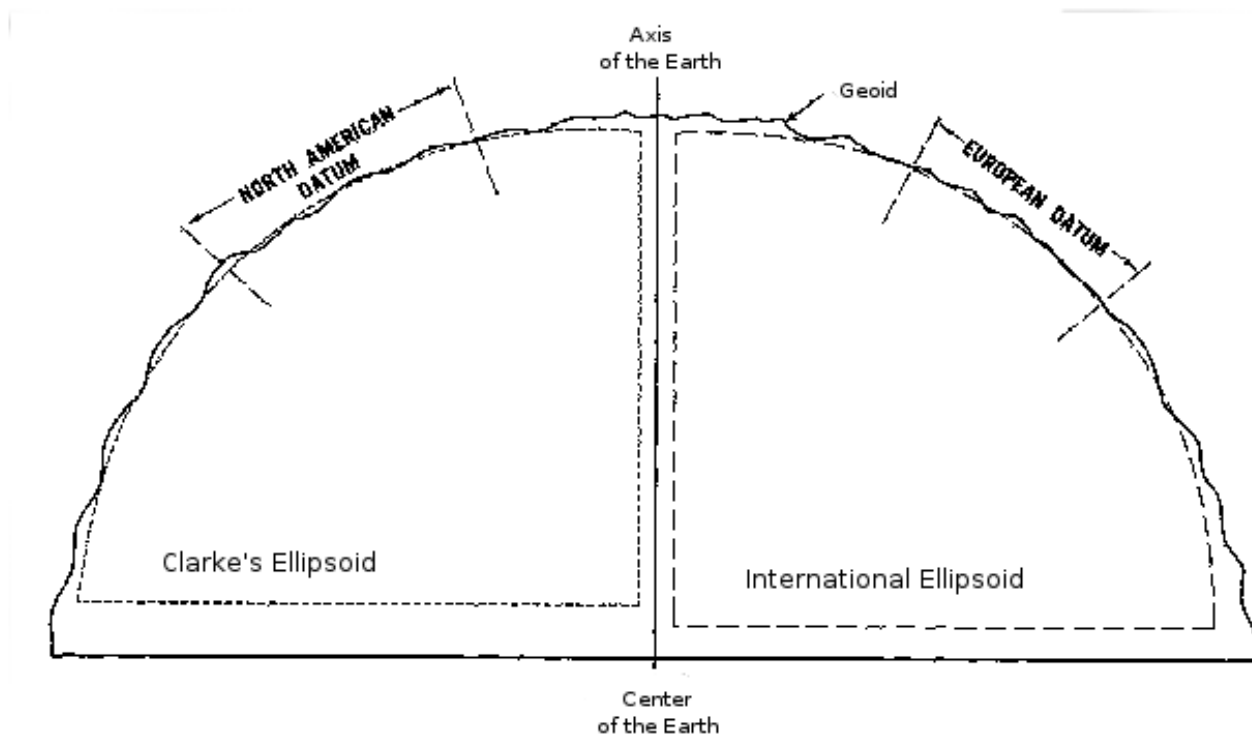


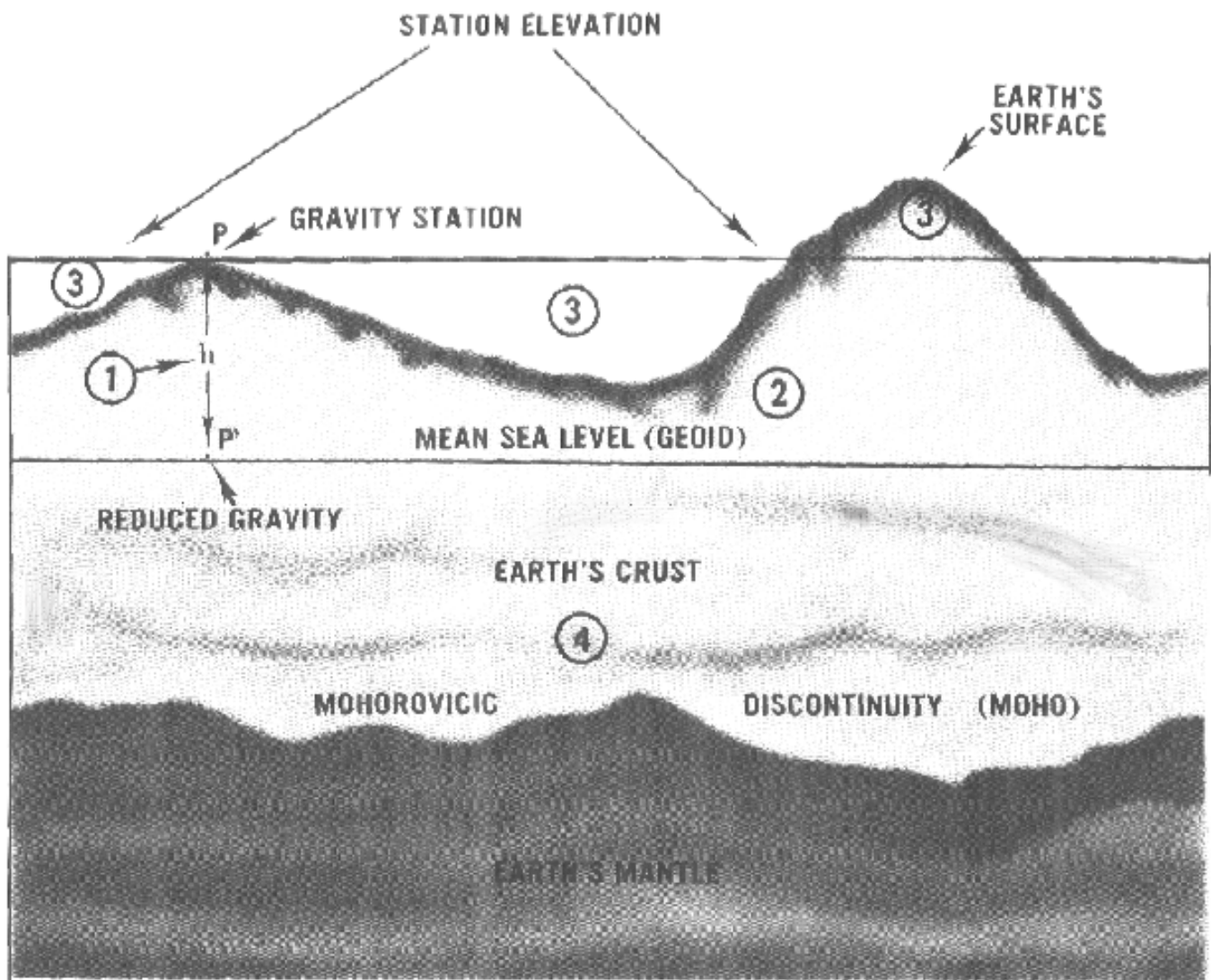
Figure 20

Figure 21



The Preferred Datums Are Relative Datums

Figure 22
**FACTORS CONSIDERED WHEN
 REDUCING OBSERVED GRAVITY**



- 1 ELEVATION, h , OF THE OBSERVATION POINT.
- 2 MASS BETWEEN STATION ELEVATION AND SEA LEVEL.
- 3 TERRAIN EFFECTS
- 4 GENERALIZED STRUCTURE OF THE EARTH'S CRUST.

The North American Datum, 1927

The first official geodetic datum in the United States was the New England Datum, adopted in 1879. It was based on surveys in the eastern and northeastern states and referenced to the Clarke 1866 Ellipsoid. Through the years this datum was extended to the south and west and in 1901, the extended network was officially designated the United States Standard Datum. The triangulation station Meades Ranch in Kansas was selected as the origin. In 1913, Canada and Mexico formally agreed to base their triangulation networks on the United States system. The datum was then renamed the North American Datum. Adjusting new surveys to fit into the network created many problems and, therefore, during the five-year period 1927-1932 all available first-order data were adjusted into a system now known as the North American 1927 Datum. The origin is still Meades Ranch and the datum is computed on the Clarke 1866 Ellipsoid which was oriented by a modified astro-geodetic method. The system not only incorporated Canada and Mexico but there are connections to the South American Datum 1969 through the West Indies and Central America. A new adjustment of the system is currently (1982) in work. The finished product will be called the North America Datum 1983 (NAD83). It will be a geocentric datum, i.e., the center of the ellipsoid will be referenced to the Earth's center of mass. The minor axis of the ellipsoid will be oriented towards a universally accepted mean pole (a mean value for the Earth's axis of rotation).

The European Datum

The initial point of this system is located at Potsdam, Germany. Numerous European national systems have been joined into a large datum based upon the International Ellipsoid which was oriented by the astro-geodetic method. The US Map Service, now known as the Defense Mapping Agency Hydrographic/Topographic Center, connected the European and African triangulation chains and filled the gap of the African arc measurement from Cairo to Cape Town. This work related the Adindan Datum in North Africa, which roughly follows the twelfth Parallel, and the Arc Datum, extending from the Equator to the Cape, to the European Datum. Through common survey stations, a datum transformation was derived between the old Russian Pulkovo 1932 and European systems. This extended the European Datum eastward to the 84th meridian. In 1946 the Pulkovo 1932 system was united with a basic Siberian network and the new datum was designated the 1942 PULKOVO SYSTEM OF SURVEY COORDINATES (Pulkovo Datum 1942). Additional ties across the Middle East connected the European with the Indian Datum.

The Tokyo Datum

The third of the initially selected datums has its origin in Tokyo. It is defined in terms of the Bessel Ellipsoid and oriented by means of a single astronomic station. With triangulation ties through Korea, the Tokyo Datum is connected with the Manchurian Datum. Unfortunately, Tokyo is situated on a steep geoid slope and the single-station orientation has resulted in large systematic geoid separations as the system is extended from its initial point.

For military distance and direction problems limited to continental areas or smaller, the large datums were satisfactory. However, while they are improvements over the limited national datums, they too have serious deficiencies which prevent them from providing the geodetic information required for intercontinental ballistic missiles.

It is true that the North American HIRAN tie permits connection of the European Datum and the North American Datum. This, however, does not completely solve the problem, for both the North American and European Datums are relative. While in each case the ellipsoid chosen is an adequate fit in the area of the origin, neither provides a good fit for the entire earth. Also, the process of connecting various datums by means of intervening datums or triangulation ties allows errors to accumulate which do not always provide agreement with newly observed data.

The South American Datum 1969

In 1965 the Working Group for the Study of the South American Datum was asked by the Committee for Geodesy of the Cartographic Commission of the Pan American Institute of Geography and History (PAIGH) to establish one consistent geodetic datum for the entire continent of South America. In June 1969, the SOUTH AMERICAN DATUM 1969 was accepted by the Commission at the IX General Assembly of PAIGH in Washington, D.C. This datum is computed on the GRS67 Ellipsoid as listed in Chapter 2. CHUA, the National datum point of Brazil was taken to be the origin. A vast amount of triangulation, HIRAN, astronomic, and satellite data were incorporated in the solution, and the datum now provides the basis for a homogeneous geodetic control system for the continent.

The Australian Geodetic Datum

The survey net of Australia consists of 161 sections which connect 101 junction points and form 58 loops. Virtually all the surveys are of the traverse type in which distances were determined by electronic measuring equipment. There are 2506 stations and the total length of the traverse is 33,100 miles. A complete readjustment of the geodetic network was made in 1966 using the GRS67 Ellipsoid as listed in Chapter 2. Tasmania has been connected and a tie to New Guinea and the Bismarck Archipelago was effected by a traverse up cape York and the USAF HIRAN network of 1965. This tie placed an additional 135 points on the Australian Geodetic Datum.

The Indian Datum

The Indian Datum has been used for India and several adjacent countries in Southeast Asia. It is computed on the Everest Ellipsoid with its origin at Kalianpur in Central India. Derived in 1830, the Everest Ellipsoid is the oldest of the ellipsoids in use and is much too small. As a result, the datum cannot be extended too far from the origin or very large geoid separations will occur. For this reason and the fact that the ties between local triangulation in Southeast Asia are typically weak, the Indian Datum is probably the least satisfactory of the major datums.

Vertical Datums

Just as horizontal surveys are referred to specific original conditions (datums), vertical surveys are also related to an initial quantity or datum. As already noted, elevations are referred to the geoid because the instruments used either for differential or trigonometric leveling are adjusted with the vertical axis coincident to the local vertical. As with horizontal datums, there are many discrepancies among vertical datums. There is never more than 2 meters variance between leveling nets based on different mean sea level datums; however, elevations in some areas are related to surfaces other than the geoid; and barometrically determined heights are usually relative.

In the European area, there are fewer vertical datum problems than in Asia and Africa. Extensive leveling work has been done in Europe and practically all of it has been referred to the same mean sea level surface. However, in Asia and Africa the situation has been different. In places there is precise leveling information available based on mean sea level. In other areas the zero elevation is an assumed elevation which sometimes has no connection to any sea level surface. China has been an extreme example of this situation where nearly all of the provinces have had an independent zero reference. There is very little reliable, recent, vertical data available for much of the area of Africa and Asia including China.

The mean sea level surface in the United States was determined using 21 tidal stations in this country and five in Canada. This vertical datum has been extended over most of the continent by first-order differential leveling. Concurrent with the new adjustment of the horizontal network, mentioned previously, is the readjustment of the vertical network. Countries of North and Central America are involved. In the conterminous United States, 110,000 kilometers of the basic network are being releveled.

Chapter 5

Physical Geodesy

Most of the discussion to this point has been concerned with geometrical geodesy. This aspect of the science is involved with strictly geometrical relationships measured in various ways-triangulation, trilateration, electronic surveys, etc.-for the purpose of deducing the size and shape of the earth and the precise location of specific positions on the earth's surface. Geometric geodesy considers the geoid through the use of astrogeodetic methods. For example, the astro-geodetic orientation method was discussed in the previous chapter.

This chapter introduces another branch of geodetic science known as physical geodesy. Physical geodesy utilizes measurements and characteristics of the earth's gravity field as well as theories regarding this field to deduce the shape of the geoid and in combination with arc measurements, the earth's size. With sufficient information regarding the earth's gravity field, it is possible to determine geoid undulations, gravimetric deflections, and the earth's flattening.

In using the earth's gravity field to determine the shape of the geoid, the acceleration of gravity is measured at or near the surface of the earth. It might be interesting to compare the acceleration measured by the gravimetrist and the acceleration experienced in an airplane. In an airplane, an acceleration is simply called a G force and is measured by a G meter. A G factor of one is used to indicate the acceleration due to the attraction of the earth and is considered a neutral condition. The gravity unit used and measured in geodesy is much smaller. A G factor of one is approximately equal to one thousand gals, a unit named after Galileo. The still smaller unit used in geodesy is the milligal (mgal) or one-thousandth part of a gal. Thus, in geodesy we are dealing with variations in acceleration equal to one millionth of a one G aircraft acceleration. The most accurate modern instruments permit measurement of acceleration changes of one hundred millionth part of the well known G factor or better.

Gravity Measurements

An attracting force, called gravitational force, operates between the earth and every object that is located within, on, or above the earth's surface. At the same time, any object within or on the earth's surface pursues a circular path as the earth rotates on its axis. If the body is on the earth's surface, it follows the circular path because it is held onto the earth's surface by the gravitational force. However, the body as it pursues a circular path exerts an outward force called the centrifugal reaction. The centrifugal reaction can be experienced by revolving an object at the end of a string.

The sum of the gravitational force and centrifugal reaction acting on a body is called gravity. Since the gravitational force is much stronger than the centrifugal reaction, gravity causes an object to have weight and, if the object is free to move, to fall with increasing speed (i.e., the body accelerates) toward the center of the earth. The acceleration experienced by that object as it moves toward the earth's center is called the acceleration of gravity, which is the quantity observed when gravity measurements are made.

Two distinctly different types of gravity measurements are made: absolute gravity measurements and relative gravity measurements. If the value of acceleration of gravity can be determined at the point of measurement directly from the data observed at that point, the gravity measurement is absolute. If only the differences in the value of the acceleration of gravity are measured between two or more points, the measurements are relative.

Absolute measurement of gravity—Until the middle of the 20th Century, virtually all absolute measurements of gravity were made using some type of pendulum apparatus. The most usual type of apparatus contained a number of pendulums that were swung in a vacuum. By measuring the period of the pendulums, the acceleration of gravity could be computed. In 1818, Kater developed the so-called reversible pendulum that had knife edge pivots at both ends. These pendulums were flipped over (reversed) during the measurements and, using this procedure, a number of important error sources were eliminated. Still, there were numerous other problems and error sources associated with pendulum measurements of absolute gravity, and the results obtained were not sufficiently accurate to meet the needs of geodetic gravimetry. Consequently, in recent years, the pendulum method has been superseded by the ballistic method which is based on timing freely falling bodies. The acceleration of gravity can be determined by measuring the time taken by a body to fall over a known distance.

In the United States, the basic falling body apparatus was developed in the early 1970's jointly by J.A. Hammond of the Air Force Geophysics Laboratory and J.E. Faller of the Joint Institute for Laboratory Astrophysics. In the so-called Hammond-Faller apparatus, a corner cube reflector falls in a vacuum while distance and time are measured continuously by a laser beam in conjunction with a photo multiplier tube. This apparatus weighed about 800 kilograms and considerably more accurate than the best absolute pendulum apparatus. Hammond recently has completed fabrication of a somewhat smaller and more accurate version of the original Hammond-Faller apparatus. The new instrument weighs about 700 kilograms when packed for shipment in nine units. Hammond's apparatus has been used to establish very accurate values for absolute gravity at a number of sites within the United States. Faller is also developing a more refined falling body apparatus. The most elaborate and probably the most accurate ballistic apparatus has been developed by Sakuma. The equipment occupies two rooms at the International Bureau of Weights and Measures at Sevres, France, and unlike the other instruments described here, is not portable. Sakuma's apparatus is a rising and falling body apparatus. A body is projected upward and allowed to rise and fall in a partial vacuum. Measurements of time and distance are made during both the rise and fall. Certain error sources cancel out when such a procedure is used. An Italian group, Istituto di Metrologia G. COLONETTI has worked with Sakuma to develop a miniaturized, portable version of Sakuma's apparatus. This portable version, generally known as THE ITALIAN APPARATUS, has been used to make very accurate measurements of absolute gravity at a number of sites in Europe and the United States since 1976. A French group, in 1977, advertised its plans to manufacture a version

of the Italian apparatus for sale.

Over the years, absolute gravity measurements have been made at only a few key sites, and these few measurements have served chiefly to establish datum and scale for relative gravity measurements. The reasons for the sparsity of absolute gravity measurements is that the necessary measuring equipment is very bulky and costly, and a single measurement requires days of painstakingly careful work under near laboratory conditions. This, however, may change in the next decade or so. As absolute equipment continues to be miniaturized and made more portable, absolute gravity measurements are becoming more commonplace.

Relative measurement of gravity—Solution of some of the problems of gravimetric geodesy requires a knowledge of the acceleration of gravity at very many points distributed uniformly over the entire surface of the earth. Since absolute gravity measurements have been too complicated and time consuming and, until recently, could not be obtained with sufficient accuracy, relative gravity measurements have been used to establish the dense network of gravity measurements needed. The earliest relative gravity measurements were made with reversible pendulums. Since the theory of relative pendulum measurements is somewhat simpler than that of absolute pendulum measurements, better accuracy was obtained for the former. However, the equipment was bulky, measurements were time consuming, and better accuracy was desired than could be obtained with pendulums. The most accurate relative pendulums to be developed were the Gulf quartz pendulum and the Cambridge invar pendulum. These two instruments were used as late as 1969.

Modern relative gravity measurements are made with small, very portable, and easily used instruments known as gravimeters (gravity meters). Using gravimeters, highly accurate relative measurements can be made at a given site, known as a gravity station, in half-an-hour or less. Modern gravimeter-type instruments were first developed in the 1930s. Although at least 28 different kinds of gravimeters have been developed, only two types have been used extensively. The LaCoste-Romberg gravimeters are used for most geodetic work today, although the Worden gravimeters have been used extensively for such work in the past.

The heart of all modern gravimeters consists of a weight suspended by a very sensitive spring. Changes in length of the spring due to changes in the acceleration of gravity as the gravimeter is moved from place to place are translated by the mechanisms of the gravimeter into DIAL READING differences which are proportional to gravity differences. The relative measurements at each gravity station consists of reading the gravimeter dial when the spring has been adjusted to a null or equilibrium position. The constant of proportionality relating dial reading differences to gravity difference is known as the calibration constant or calibration factor. The dial reading of the gravimeter at each site is multiplied by the calibration factor to obtain a gravity value. Each instrument has a unique calibration factor which must be determined empirically. This is done by the manufacturer. Many gravimeter users redetermine and periodically check the calibration factor by taking dial readings over a so-called calibration line. A calibration line is a series of well described monumented, reoccupied sites where the acceleration of gravity has been determined very accurately and over which the value of gravity varies significantly. By comparing the dial readings to the known gravity values at points along the calibration line, the calibration factor can be computed. In this manner, the scale of relative gravity surveys is controlled by the calibration factor. For the most precise work, it cannot be assumed that the calibration factor is constant, and more complicated

calibration procedures must be used.

There are two other important considerations when relative gravity measurements are made: drift and base station connections. Gravimeter drift is a phenomenon related to certain instrumental instabilities that cause the dial reading to change slowly with time even when the acceleration of gravity remains constant. Although the amount of drift is very small for most modern gravimeters, it still must be determined and eliminated from the relative gravity measurements. There are several observational techniques commonly used, but all involve some kind of looping back to previously measured points such that the relative gravity survey is continuously closed back on itself. The difference in dial readings (converted to gravity) between two successive occupations of the same site is the drift of the gravimeter over the interval of time which has elapsed between the two occupations. Drift is often considered to occur at a constant rate. Consequently, the time at which each measurement is made is recorded along with the dial reading. After each closure in the survey, the total drift is determined and allocated back through the loop proportionally to the time elapsed between each measurement. For the most accurate work, drift cannot be assumed to be constant, and more complicated techniques should be employed. For the most accurate gravity surveys, several gravimeters are used simultaneously, and a least squares adjustment is applied to compensate for imprecise drift and calibration determinations.

Since relative gravity surveys can determine only differences in gravity from point to point, every relative gravity survey must include measurements at one or more reoccupiable points where acceleration of gravity is known. Such points are called base stations. Then all gravity difference measurements are computed with respect to the known gravity value at the base station. Hence, tying a relative gravity survey to a base station establishes the GRAVITY DATUM of that survey. The earliest GRAVITY DATUM was the so-called Potsdam System. In 1906, reversible pendulums were used to measure absolute gravity at Potsdam, Germany. The value measured at this time was later adopted as the initial point for the Potsdam gravity reference system. Using relative measurements, several points on each continent were connected to Potsdam, and these served as the fundamental base stations for many relative gravity surveys.

The Potsdam system, however, was found to be in error and, in 1971, was replaced by the International Gravity Standardization Net 1971 (IGSN71). The IGSN contains 1854 reoccupiable stations distributed worldwide. The acceleration of gravity at each point was determined by a least squares adjustment that included a number of absolute gravity measurements and a multitude of relative gravity measurements that interconnected all stations. The IGSN71 established the basic GRAVITY DATUM for today's relative gravity surveys. National base networks exist in many countries. These national base networks usually are connected to the IGSN71 and contain reoccupiable stations located at convenient sites throughout the country concerned. For example, the US gravity base network contains approximately 50 stations. The national base networks usually are established by precise gravimeter measurements. In the future, some national nets will be tied directly to high accuracy absolute gravity measurements instead of the IGSN71. Actually any reoccupiable gravity station which has been tied to the IGSN71 or a national base network can be used as a base station for a new relative gravity survey.

Gravity measurement at sea—So far, only procedures used to measure gravity on land areas have been discussed. Gravity measurements, however, also are made at sea. The earliest measurements at

sea were made by F.A. Vening Meinesz who, in 1927, installed a pendulum apparatus in a submarine. The submarine pendulum gravity measurements of Vening Meinesz are mainly of historical interest today. The first gravimeters installed in surface ships appeared during the 1950's. These early ocean surface gravity measurements were only of modest accuracy and, again, now are mainly of historical value. Reasonably accurate measurements from gravimeters on surface ships date only from the late-1960's. Instruments used include LaCoste-Romberg S Meter, Askania Meter, Bell Meter, and the Vibrating String Gravimeter. All of these meters are compensated to minimize the effects of oscillatory motion of the ship due to ocean surface waves. The effects are also eliminated or averaged out by computational techniques. A big problem with ocean surface measurements is that the forward motion of the ship adds a centrifugal reaction component to measured gravity which must be eliminated by the so-called Eotvos correction. Therefore, the ship's velocity and heading, as well as the ship's position, must be known accurately. Near shore, shore based electronic positioning/navigation systems (such as LORAN) are used. In the deep ocean, satellite navigation and inertial systems must be used.

Gravity measurement in the air-Problems in airborne gravity measurements are similar to those encountered for surface ships. The position, velocity, and heading of the aircraft must be known accurately. Because of the higher aircraft speeds, the Eotvos correction is much larger for airborne measurements than for surface ship measurements. It also is very difficult to compensate for spurious aircraft accelerations. In addition, reduction of the gravity value from aircraft altitude to an equivalent surface value is a problem that has not yet been solved satisfactorily. In 1959, the US Air Force was instrumental in developing a gravimeter suitable for use aboard a fixed wing KC-135 aircraft. This gravimeter measured average gravity values along the path of the aircraft reasonably well, but could not produce point gravity values. The program was canceled in the late 1960's. Recent suggestions have been made that fixed wing airborne measurements should be restudied. Today, there is an operational airborne gravity measurement system aboard a S-61 helicopter. The gravity measurement system aboard the helicopter uses a LaCoste-Romberg S Meter to sense gravity and a Del Norte transponder electronic navigation system to establish aircraft position, velocity, and heading. The operator of the system, Carson Helicopters, advertises that it is possible to do 3,000 to 10,000 miles of gravity survey per month with reasonably good accuracy.

Gravity Anomalies

Gravity measurements provide values for the acceleration of gravity at points located on the physical surface of the earth. Before these measurements can be used for most geodetic purposes, they must be converted into gravity anomalies.

A gravity anomaly is the difference between a gravity measurement that has been reduced to sea level and normal gravity. Normal gravity, used to compute gravity anomalies, is a theoretical value representing the acceleration of gravity that would be generated by a uniform ellipsoidal earth. By assuming the earth to be a regular surface without mountains or oceans, having no variations in rock densities or in the thickness of the crust, a theoretical value of gravity can be computed for any point by a simple mathematical formula. The theoretical value of gravity at a point on the ellipsoid's surface depends both on the size and shape of the ellipsoid and on a value, computed from observational data, which is accepted as the theoretical value of gravity at the equator. It varies only with the latitude of the observation point if the figure of the earth is taken as an

ellipsoid of revolution. While there are several formulas for computing Theoretical Gravity, the one most commonly used is the Gravity Formula 1967 which is based on the Reference Ellipsoid 1967, the parameters of which correspond to the Geodetic Reference System 1967. A simple formula will convert 1967 theoretical (or normal) gravity to the newer more accurate 1980 system, GRS80 (Chapter 2). A number of different procedures can be used to reduce a gravity measurement from the physical surface to sea level and, depending upon which procedure is used, different types of gravity anomalies result. The most common type of gravity anomaly used for geodetic applications is the so-called free-air gravity anomaly.

Since there are variations in the densities of the earth's crustal materials as well as terrain variations, the observed gravity of the earth varies irregularly from point to point. As a result the surface known as the geoid is an irregular figure. As explained earlier, the distance between the mathematical ellipsoid and the actual geoid is called the undulation of the geoid. Since the geoid is so irregular, geoid undulations cannot be computed directly but must be determined point by point. Gravity measurements provide the observational data for these computations.

The word, anomaly, as used in geodesy refers to a deviation from the normal and can be used either for a single point or to describe a regional or area effect. To represent an area of the earth's surface, the gravity measurements (anomalies) within the area are averaged by some specialized technique to produce a mean value. This may be done for areas of different size such as the one degree by one degree square. To make use of the anomalies, the observed gravity must be reduced to a common frame of reference, the geoid-mean sea level. The reductions may take into account the elevation above (or below) sea level and can account for the mass between the point and sea level, the surrounding terrain, and the assumed structure of the earth's crust.

Undulation and Deflections by the Gravimetric Method

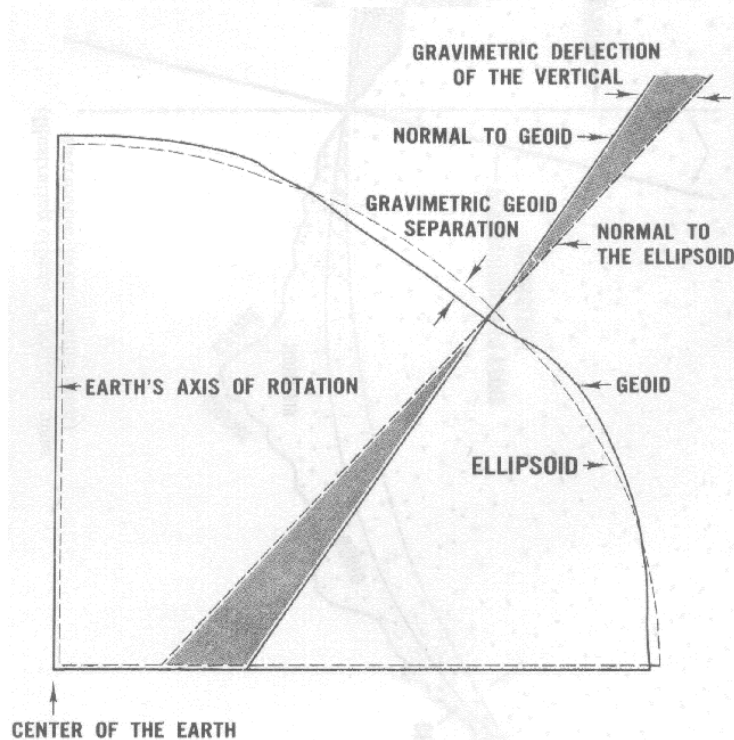
The method providing the basis from which the undulations of the geoid may be determined from gravity data was published in 1849 by a British scientist, Sir George Gabriel Stokes. However, the lack of observed gravity data prevented its application until recent years. In 1928, the Dutch scientist, Vening Meinesz, developed the formulas by which the gravimetric deflection of the vertical (Figure 23) can be computed.

The computation of the undulations of the geoid and the deflections of the vertical require extensive gravity observations. The areas immediately surrounding the computation point require a dense coverage of gravity observations and detailed data must be obtained out to distances of about 500 miles. A less dense network is required for the remaining portion of the earth. While the observational requirements for these computations appear enormous, the results well justify the necessary survey work.

Figure 24 shows how the mass surplus of the mountains and the mass deficiency of the oceans cause the deflections of the vertical and the undulations of the geoid. A mountain mass attracts the plumb line from the normal of the ellipsoid. Likewise the mass deficiency of the ocean does not attract the plumb line. These effects of the mass anomalies contribute to the deflection of the vertical. However, both deflections of the vertical and undulation

Figure 23

PRODUCTS OF THE GRAVIMETRIC METHOD



values result from density variations throughout the earth. In the area of mass surplus, the observed gravity (reduced to sea level considering only the elevation) is generally greater than the theoretical value, and the anomalies are positive. In the areas of mass deficiency, the observed (reduced in the same manner) is generally smaller than the theoretical value and the anomalies are negative.

The deflections and undulations computed with sufficient gravity information are considered absolute values referred to an earth-centered reference ellipsoid. In other words, the axis of rotation for the ellipsoid passes through the earth's center of gravity.

Effective use of the gravimetric method is dependent only on the availability of anomalies in sufficient quantity to achieve the accuracy desired. Successful use of Stoke's integral and Vening Meinesz formulas depends on a good knowledge of gravity anomalies in the immediate vicinity of the point under consideration and a general knowledge of anomalies for the entire earth.

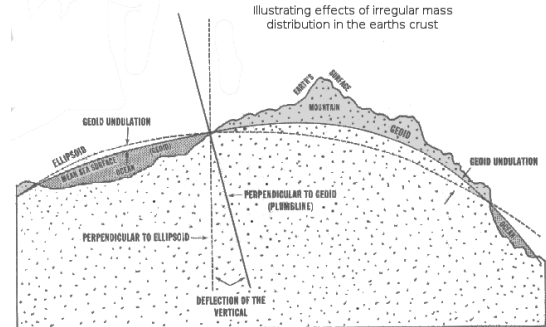
Geophysical Gravity Predictions

Measured gravity coverage is by no means complete. There are many large regions on the continents where gravity measurements are lacking or available only in small quantities. Gravity

Figure 24

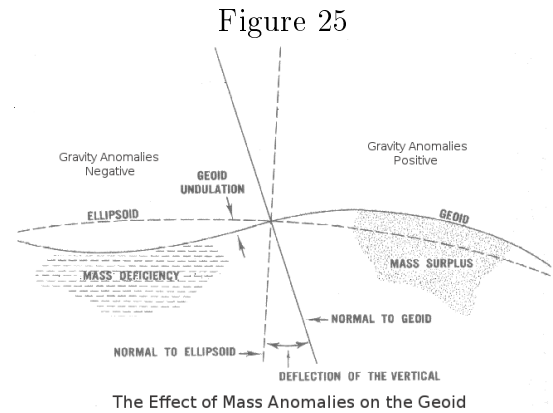
Geoid Ellipsoid Relationships

Illustrating effects of irregular mass distribution in the earth's crust



data for ocean areas has always been sparse, however, Satellite Altimetry (Chapter 7) has overcome this deficiency. In regions where an insufficient number of gravity measurements exists, some other approach must be used to obtain or predict the mean gravity anomalies for the areas.

Correlations exist between variations in the gravity anomaly field and corresponding variations in geological, crustal, and upper mantle structure, regional and local topography and various other types of related geophysical data. In many areas where gravity information is sparse or missing, geological and geophysical data is available. Therefore, the various prediction methods take into account the actual geological and geophysical cause of gravity anomalies to predict the magnitude of the anomalies.



Chapter 6

Satellite Geodesy

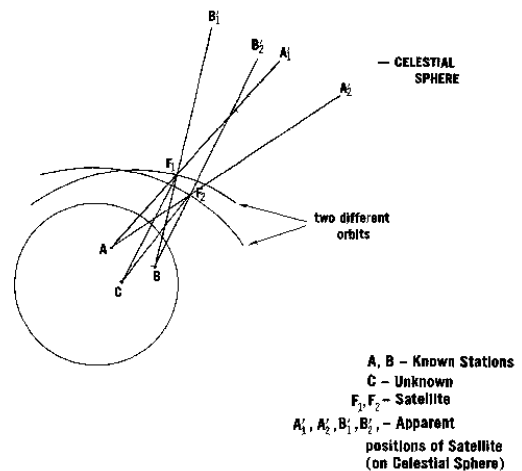
Scientific papers advocating the use of satellites for geodetic purposes were published as early as 1956. Geodetic applications were outlined by the Smithsonian Astrophysical Observatory for data obtained from Project Vanguard during the 1958-59 International Geophysical Year. Many techniques and a great deal of knowledge were ultimately derived from this project. With this information, the constant growth of space technology, the development of electronic distance measuring devices, and the perfection of electronic data processing equipment, satellites specifically equipped for geodetic purposes have been developed, launched, observed and the data utilized.

The first real geodetic satellite was ANNA-1B launched in 1962. Project ANNA was a truly cooperative effort involving the Department of Defense (DoD), the National Aeronautics and Space Administration (NASA), and other civil agencies. Several observational systems were developed and improved during ANNA. These systems included geodetic cameras, electronic ranging and Doppler. Knowledge gained from Project ANNA was also useful in the development of Laser ranging systems.

Observational Systems

Two basic systems have been used for obtaining geodetic information from artificial earth satellites-optical and electronic. These systems have made it possible to perform various geodetic measurements to relate known or unknown positions to the earth's center, to relate unknown positions to existing triangulation networks, and to relate the triangulation networks to each other. Important parameters of the earth's gravitational field and values for the earth's flattening have also been obtained.

Figure 26



Historical Systems

Optical tracking with the BC-4, PC-1000, MOTS, or Baker Nunn cameras consisted of photographic observations of a satellite, or flashing light on the satellite, against a background of stars. The stars, whose positions were accurately determined, provided a framework on the photographic plate or film for a determination of precise directions from camera station to satellite. Geodetic positioning work with cameras was usually performed with one camera observing simultaneously with one or more other cameras. Camera systems are weather dependent and that is one major reason why they are little used today. Laser systems discussed later, are also weather dependent but their extreme accuracy justifies their use and development.

The U.S. Army developed the SECOR (Sequential Collation of Range) system and the first SECOR transponder was orbited on ANNA-1B in 1962. The SECOR system continued in use through 1970. The system operated on the principle that an electromagnetic wave propagated through space undergoes a phase shift proportional to the distance traveled. A ground station transmitted a phase modulated signal which was received by the satellite-borne transponder and returned to the ground. The phase shift experienced by the signal during the round trip from ground to satellite and back to ground was measured electronically at the ground station which provided as its output a digitized representation of range.

Doppler

A geodetic satellite can also carry electronic signaling equipment to produce the Doppler effect which can be used for geodetic purposes. The Doppler observational system was derived from the fact that while a satellite transmitter sends a continuous unmodulated wave at a fixed frequency, the received signal at the tracking stations exhibits a shift in frequency due to the relative velocity of the satellite and observing station. A similar phenomenon may be observed with sound waves, as the source of the sound approaches and recedes from the observer. For example, the pitch of a train whistle apparently changes as the train approaches and recedes from the observer. Although the sound waves travel at a constant rate—approximately 1080 feet per second at sea level—they become crowded together as the source approaches the observer, the wave lengths become shorter, and the pitch increases. The opposite effect takes place as the source moves away.

In Figure 27, the radio source is assumed to be moving in a straight line at a specified speed. The angle ϕ is used to indicate the distance between the position of the satellite at P (time t) and the point P_m (time t_m) where it will be closest to the observer. If the Doppler shift were expressed in terms of ϕ , as ϕ decreases, the Doppler shift decreases to zero. As the satellite recedes, the received frequency decreases and the shift increases until the transmitter is out of range. This received frequency is actually a function of the transmitted frequency, velocity of propagation, and the rate of change of the slant range between the satellite and station. From observations at one station, the satellite period, time and distance of its closest approach, and its relative velocity can be determined. If observations are made from three or more stations, the orbital parameters may be derived.

Of all the satellite observation systems used so far, Doppler tracking has been the most fruitful. Reasons for this are: it is passive, not requiring any interrogation or directionally sensitive antennae at the receiver; the data obtained (Doppler counts) are in digital form; the radio frequencies used

permit all weather day and night tracking; and accuracies achieved have steadily improved. Much effort has gone into identifying and eliminating the sources of error in Doppler work. Automatic portable receiving equipment is available from several suppliers. The GEOCEIVER (geodetic receiver) has played an important role in DoD observational programs. Since 1967 there have been at least four operational satellites continuously available for Doppler positioning. The National Geodetic Survey (NGS) will be using results from many Doppler stations within the United States and its possessions in support of the readjustment of the North American horizontal datum.

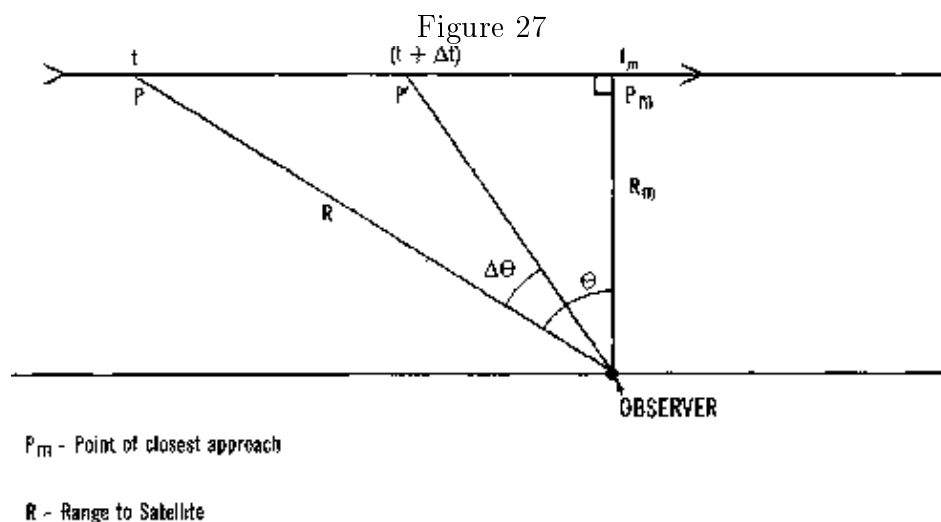
Laser

The laser has been adapted to measuring distances over the earth's surface and for computing ranges from earth stations to satellites and the lunar surface. The laser instrument is pointed to a target and then activated by a clock at the appropriate time. The laser beam is reflected at the

target by special reflectors and the returning light is detected photoelectrically, and its time of flight measured to yield range data. The laser transmitter is mounted adjacent to some type of telescope or optical device used for receiving the reflected laser beam.

In satellite laser ranging, the interval between the outgoing and returning pulse from the satellite is measured very accurately and then transformed into a range measurement which is

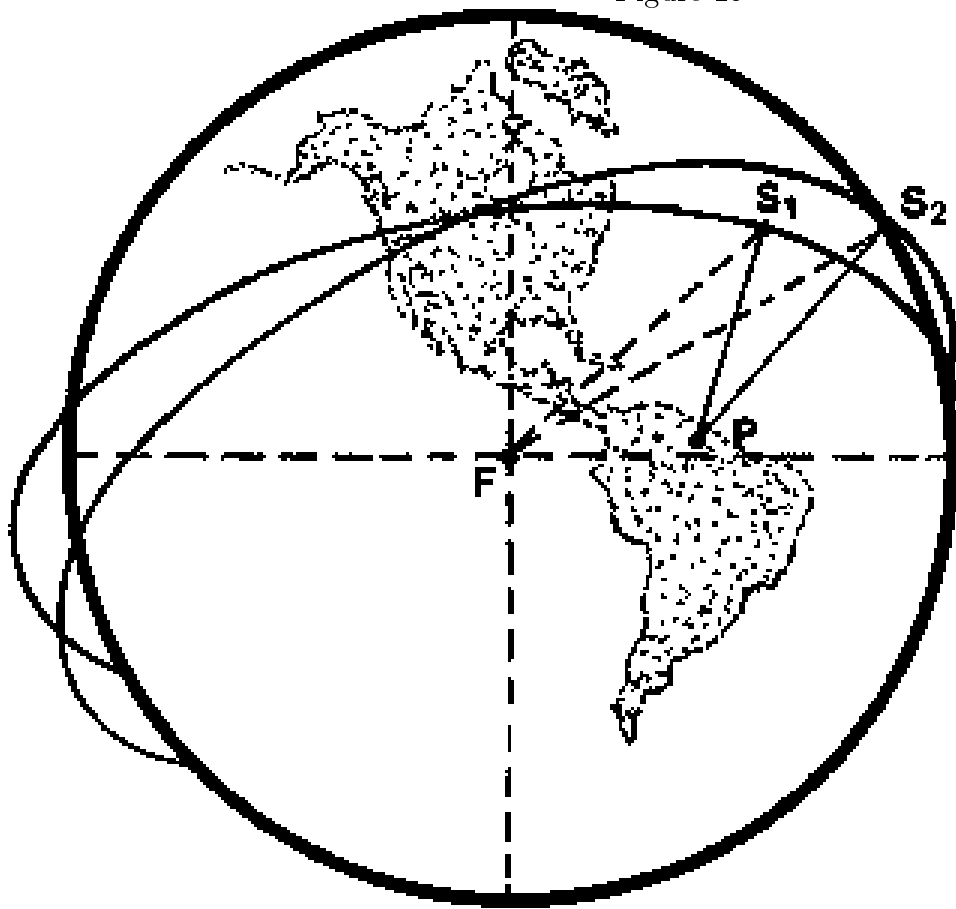
corrected for atmospheric refraction. Laser ranging is possible even when the satellite is in the earth's shadow and during daylight hours.



Doppler Shift

Simultaneous laser ranging to a near-earth satellite from two sites is used to determine the coordinates of one laser site relative to the fixed position of the other site and simultaneously the inter-site distance. NASA has used laser tracking since 1972 to measure the distance between points in North America. They have been testing the accuracy of laser tracking in measuring the crustal motion between points on opposite sides of the San Andreas fault and plan to make repeated measurements of baselines across the fault over a number of years. Simultaneous laser tracking has also been achieved between an east coast site and Bermuda enabling a determination of the Bermuda site's relative location (North American Datum) and the baseline between the two sites.

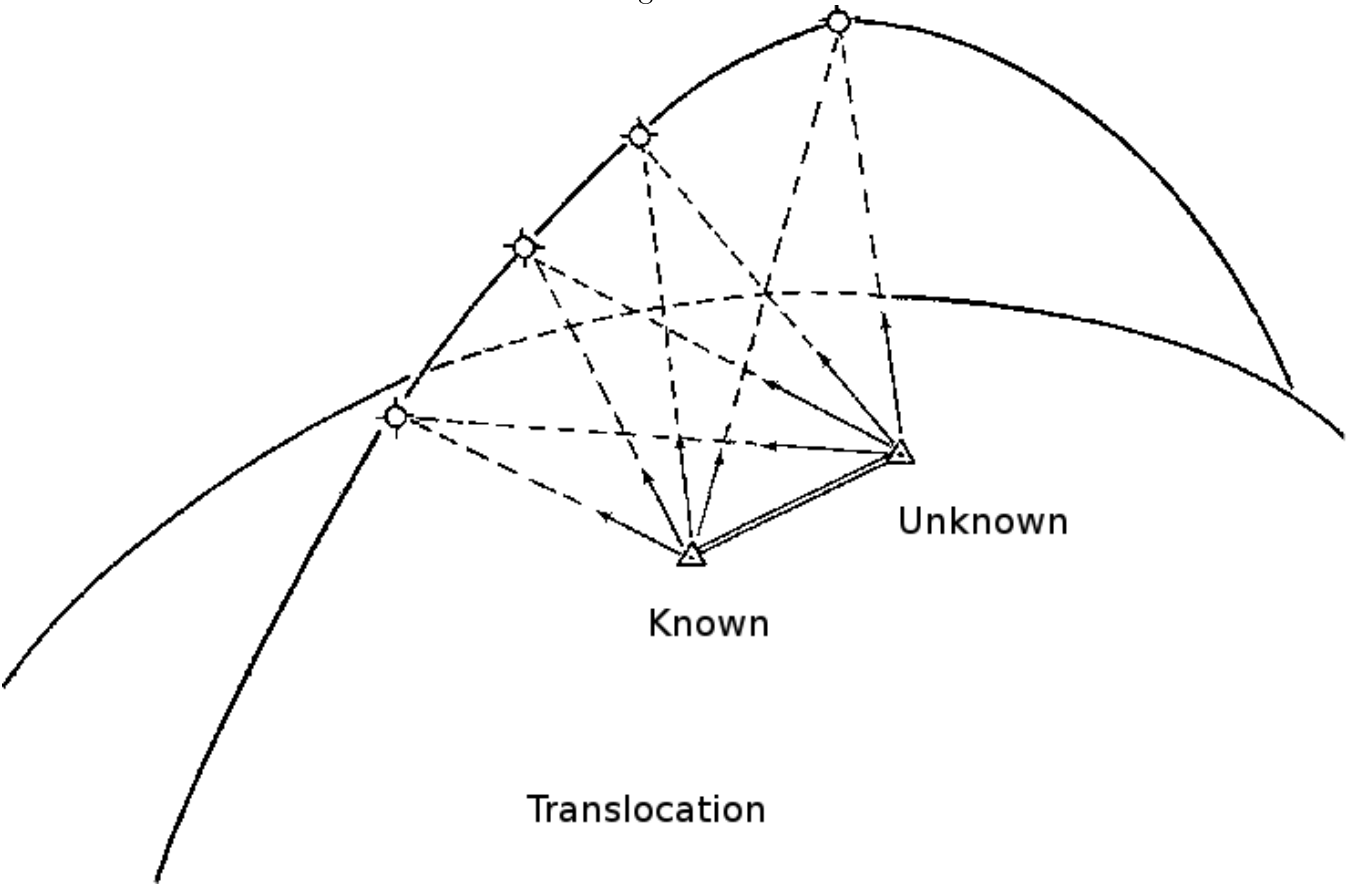
Figure 28



Point Positioning

P -OBSERVATION STATION
F -CENTER of the EARTH
S₁- S₂ -POSITIONS of SATELLITE
FROM TWO DIFFERENT
PASSES.

Figure 29



Laser ranging data has been incorporated into the development of world geodetic systems by the Smithsonian Astrophysical Observatory (SAO) and the Department of Defense (DoD). NASA has also included laser data in their development of gravitational models. Laser data is also being used for polar motion and earth rotation studies.

Doppler Satellite Surveys

Doppler satellite surveying is a method of determining positions of points on the earth's surface by observing the Doppler shift of radio transmissions from satellites of the U.S. Navy Navigation Satellite System (NNSS). NNSS was developed for the Navy as a worldwide all-weather navigation system, and provides position fixes at time intervals of 2 hours or less. Observations of these satellites began in earnest in 1971, when portable tracking receivers became available to establish precise positions on the earth's surface. Since then, the application of Doppler satellite surveying techniques expanded, until today they are used worldwide.

System Description

The NNSS satellites currently operational are in circular polar orbits approximately 1000 km above the earth's surface. The time to complete an orbit (orbital period) is approximately 105 minutes. The earth's rotation causes a satellite to cross the Equator on each revolution approximately 26° in longitude west of the previous crossing. Each satellite transmits what is referred to as the BROADCAST EPHEMERIS, which describes the satellite's position in space. The ephemeris is a predicted orbit based on Doppler observations previously acquired by four tracking stations located in the United States.

Doppler satellite observations, reduced using the broadcast ephemeris, yield point positions with sufficient accuracy to satisfy the requirements of many NNSS satellite users. There exists a more accurate PRECISE ephemeris which is generated for selected NNSS satellites from Doppler observations acquired by a tracking network (called TRANET) composed of 15 to 20 stations. The positioning accuracies for single station observations are greatly improved when using the precise ephemeris. These ephemerides are computed by the Defense Mapping Agency (DMA) to support geodetic applications requiring point positions of one-meter accuracy.

Modes of Observation

Points on the earth's surface can be positioned by various modes (point positioning, simultaneous point positioning, translocation, semishort-arc or short-arc) using either the broadcast ephemeris or the precise ephemeris.

In the point-positioning mode, multiple passes collected with a single Doppler receiver are used with an ephemeris to determine an independent station position in geocentric coordinates (X,Y,Z) referenced to the Earth-centered satellite coordinate system. The geocentric coordinates can also be expressed in geodetic coordinates (latitude, longitude, and height above ellipsoid). Doppler positions determined with the precise ephemeris can be directly transformed to the World Geodetic System 1972 (WGS72) discussed in Chapter 8.

When employing the point-positioning mode as the Doppler surveying technique, the user may wish to transform the Doppler satellite-derived geocentric coordinates to the local geodetic system. The process of deriving the coordinate shifts and transforming the Doppler position requires a thorough understanding of datum transformation concepts and procedures. Occupation of a station with known local geodetic coordinates allows the transformation parameters to be derived for subsequent use with Doppler positions that were not established on a station tied to the local geodetic system.

In the simultaneous point-positioning mode, stations are simultaneously occupied in figures of two or more. The data are independently reduced as in the point-positioning mode and differenced to form relative positions. The simultaneous observations are performed during a common time period, but do not necessarily include common satellite passes.

In the translocation mode, observations are simultaneously collected, usually at two stations. This mode is employed to obtain very accurate relative positions, even if the precise ephemerides are not available. When the broadcast ephemeris is used, statistical correlation performed during data reduction improves the accuracy of the positioning. The principal error sources affecting an individual satellite position fix are the ephemeris errors and refraction errors. Improved compensation for these errors is possible when the same signal is received at separate sites. The maximum spacing between sites is generally limited to approximately 500 km (or less if comparability with existing control is to be maintained) so that desirable portions of satellite passes can be tracked simultaneously. During processing, enforcement of simultaneity of data points is optional. When simultaneity is enforced, it is generally referred to as rigorous translocation.

The short-arc and semishort-arc modes allow for small adjustments in the orbit instead of holding the satellite ephemeris fixed, as is done for the other methods. The translocation method, for example, assumes that orbit errors affect positioning of all sites in the same way, whereas the short-arc technique adjusts the reference orbit while simultaneously solving for positions. In short-arc processing, six orbital parameters are allowed to adjust. The method becomes semishort-arc processing when one to five parameters are adjustable.

Short-arc, semishort-arc, translocation and simultaneous point positioning are all used for relative positioning. In a survey scheme, one or more base station positions are generally held fixed. These fixed positions usually have a known position on a given geodetic datum. While data are being collected at the fixed station(s), one or more additional receivers are circulated among the various unknown stations comprising the net. The data collected simultaneously are subjected to post processing to determine a position relative to the base stations.

Although point positioning is the least accurate mode to use when only the broadcast ephemeris is available, this technique can be performed with a single receiver and fairly simple computations. Translocation and short-arc techniques are the most accurate modes when the precise ephemeris is not available, but for these modes, field and computational procedures are more complex.

Because NNSS satellites are capable of providing the broadcast ephemeris, it may be more economical to use the broadcast ephemeris rather than the precise ephemeris to meet the survey requirements. This can reduce significantly the period of occupation for a station. Data processing can also be performed more quickly because no time is lost waiting for post tracking orbital data

in order to generate the precise ephemerides.

Harmonic Analysis of Orbital Data

A great deal of study has been done regarding the effect of the earth's gravitational attraction on satellite motion. The fact that there are a number of perturbing factors has already been mentioned—the uneven distribution of the earth's mass, the oblateness of the earth, atmospheric drag, the effects of the planets, sun and moon, and electromagnetic effects. The perturbations are measured by observing the position of the satellite in orbit around the earth. As observational data accumulates, orbital parameters become more precisely defined and reliable earth-centered positioning becomes available. An analysis of orbital data can also be used to develop an expression of the earth's external gravity field for a better interpretation of the shape of the geoidal surface through spherical harmonics. (The Appendix contains a discussion of spherical harmonics.) Although a complete analysis of orbital data requires consideration of all perturbing effects, the earth itself is the only perturbing body of major consequence in the study of near-earth satellite motion. The effects of the sun, moon, and atmosphere are removed so that only the effects of earth's shape and uneven mass distribution remain.

The uneven distribution of the earth's mass causes the force of gravity to vary from point to point on the surface and in external space. While force of gravity is measured at points on the surface with highly sensitive instruments, mathematical procedures are required to analyze orbital perturbations and to express the gravitational potential. The gravitational potential may be explained in terms of potential surfaces—surface to which the force of gravity is always perpendicular. If the earth were a perfect non-rotating sphere with homogeneous mass distribution, the potential surface would be spherical in shape. The fact that the earth is shaped more like an ellipsoid than a sphere causes the potential surface to be shaped more like an ellipsoid. Actually, the earth is neither spherical nor ellipsoidal. The potential surface bulges where there is excessive mass and it is depressed in areas of mass deficiency. The undulating surface described earlier as the geoid is a potential surface of the real earth. The diagram in Figure 31 illustrates the three surfaces just discussed.

The most convenient way to express the gravitational potential is in terms of a series of spherical harmonics mentioned above. The coefficients of the various harmonic terms are functions of the various orbital perturbations. A few are directly related to the shape of the earth and the remainder to the uneven distribution.

While it is possible to derive harmonic coefficients from observed gravity, the method is limited due to the lack of high quality worldwide gravity coverage. The computation of coefficients from satellite data also has its limitations. There are many coefficients that are not well defined from tracking data due to the small magnitude of the orbital perturbations at geodetic satellite altitudes. In addition, satellites orbiting at different inclinations are needed to reduce the correlation between the computed coefficients. For best results, the current practice is to combine tracking data with available surface-gravity data when solving for the spherical harmonic coefficients of the earth's gravitational field.

Figure 30

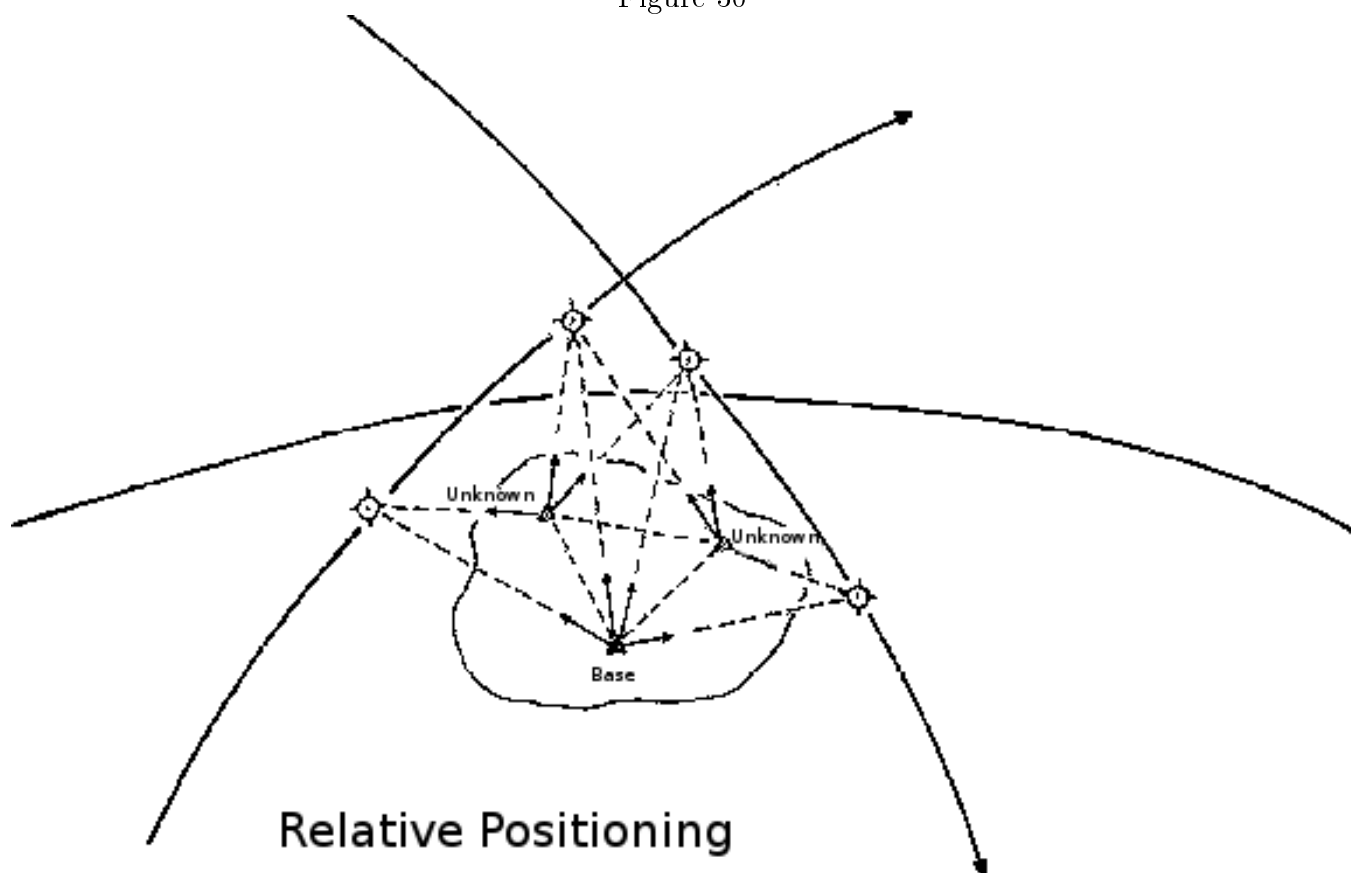
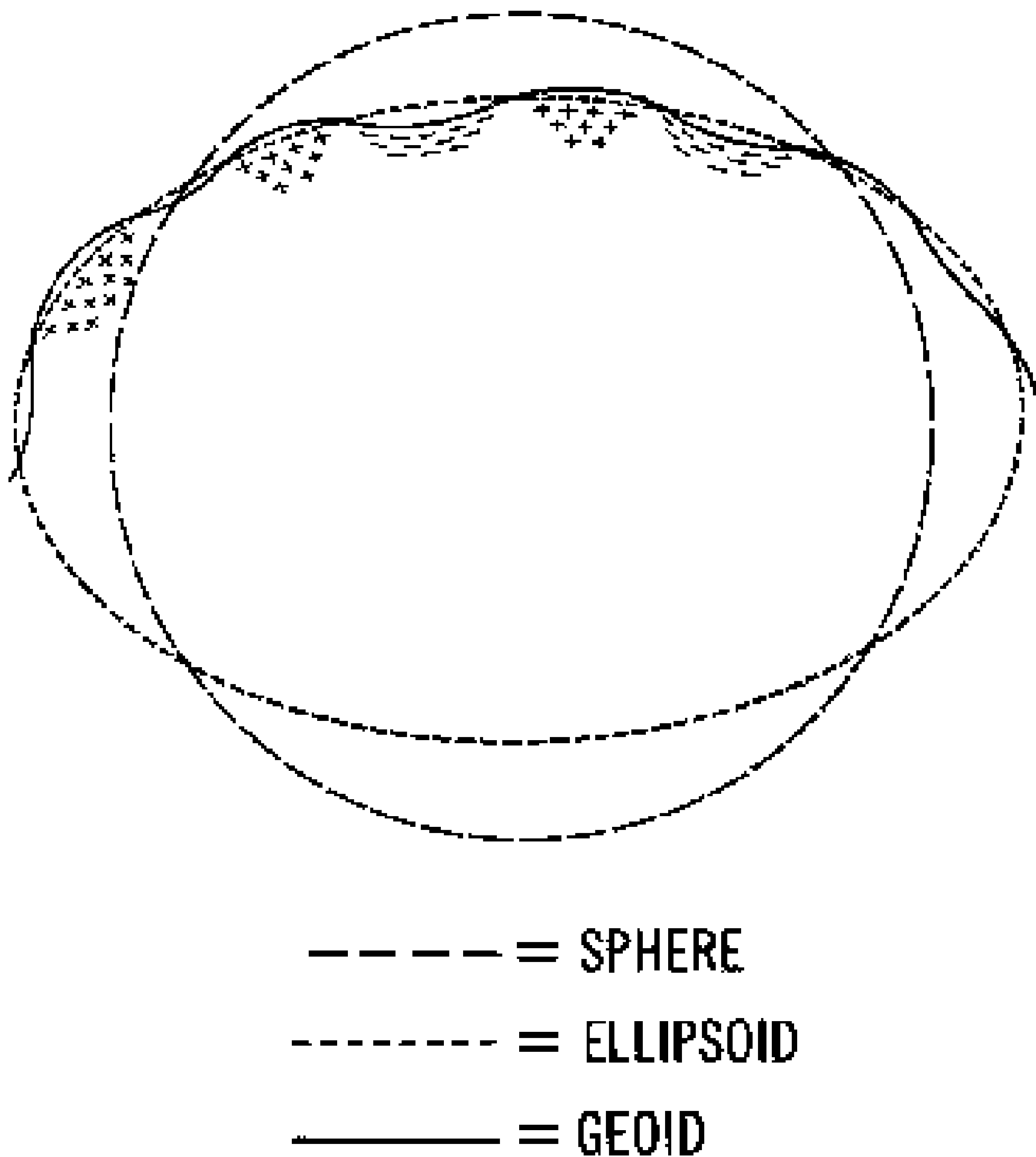


Figure 31

Potential Surfaces



Chapter 7

Other Developments in Geodesy

Many pertinent developments have occurred since the last revision of this publication in February 1968. This chapter provides a place for subjects of interest that are not discussed in the other chapters. Some of the areas of new geodetic developments are: satellite laser ranging (Chapter 6), lunar laser ranging, very long baseline interferometry, satellite radar altimetry, the NAVSTAR Global Positioning System, satellite-to-satellite tracking, and inertial surveying.

Lunar Laser Ranging

In July 1969, the first men to set foot on the moon performed a number of tasks of scientific importance. Among these tasks was the deployment of a rack structure carrying an array of 100 fused silica retro-reflectors designed to return some of the light of a pulsed laser beam to the telescope to which the laser equipment is coupled. These retro-reflectors are a part of an Apollo experiment called LURE (Lunar Laser Ranging Experiment). LURE was designed to obtain extremely accurate measurements of the range from known locations on the surface of the earth to the lunar based retro-reflectors and enable the improvement of our knowledge of the rotation of the earth and the moon about their center of mass and the moon's libration and motion about the earth.

Very Long Baseline Interferometry

Observations of extragalactic radio sources such as quasars, can provide the geodetic information to determine the vector separations between the antennas of two widely separated radio telescopes. The components of the vector are its length and direction. To accomplish this, it is necessary to measure very accurately the difference in the time of arrival, recorded at the two antennas, of a particular wavefront from a given (point) source of radio radiation. The phenomena called interference, in Very Long Baseline Interferometry (VLBI), is produced by electronically superimposing the recorded signals to produce a resultant disturbance or INTERFERENCE pattern. The theoretical expression for the relative phase delay shows it to be a function of the source direction, the antenna locations, the relative clock error between the two sites, the time of day, the model atmosphere employed, the earth's tidal parameters, the radio frequency at which the observation is made, etc. Proper account must also be taken of the earth's rotation. Two of the main limiting factors in the VLBI technique are clock stability and atmospheric variations. A major goal of VLBI is to reduce the uncertainty in intercontinental baselines to the centimeter level.

VLBI derived baselines have already contributed scale information to the development of the DoD World Geodetic System in 1972. Baselines accurate to the centimeter level would function as standards of comparison for future world systems. Other applications of VLBI include the determination of polar motion, variations in the earth's rotation, and the monitoring of motions of the major plates that compose the earth's crust.

Satellite Radar Altimetry

The development of orbiting space satellites from which microwave remote sensing of the earth can be achieved has provided a new instrument to the geodesist which measures directly the shape of the geoid in the ocean areas. The satellite altimeter consists of a downward ranging radar which measures the time delay from the transmission to the reception of a pulse of energy.

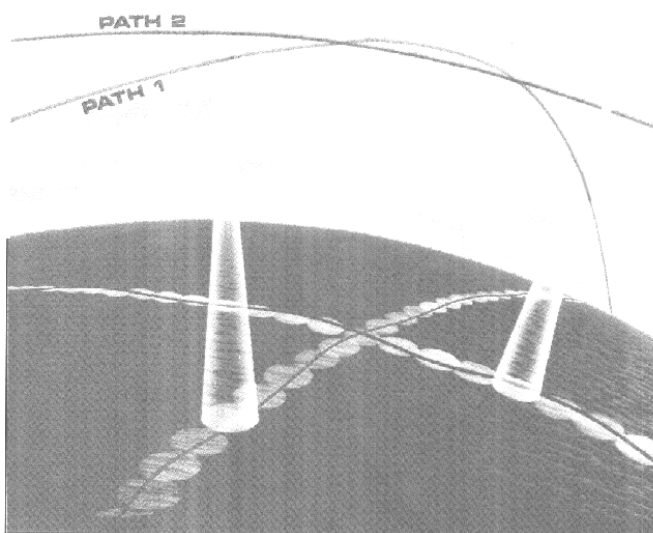
The observed one-way distance from the transmitting antenna to the surface is equal to one-half the product of the time delay and the speed of light. From this distance or height, the local surface effects such as tides, winds and currents are removed to obtain the satellite height (h) above the geoid. With a precise Doppler ephemeris available for the satellite, the radius R_{SAT} to the satellite, determined for the time of each observation, along with the radius R to the ellipsoid are readily at hand. It is then possible to compute the geoid height (N) by subtracting the radius R and the satellite height h from R_{SAT} .

The Skylab spacecraft, launched in 1973, provided the first opportunity for satellite based radar altimetry. It was basically a research mission for which data was obtained for the designing of future altimeters. The GEOS-3 altimeter which incorporates many of the design features that were tested in the Skylab altimeter was launched in 1975 and provided geoid measurements over the water areas of the earth from 65° N to 65° S. The SEASAT altimeter which was a more sophisticated instrument with greater measurement capabilities was launched in June 1978 and added data from 72° N to 72° S.

The NAVSTAR Global Positioning System

Scientists, engineers, and planners have been tasked with making comprehensive studies of currently available navigation systems as part of an effort to devise a system capable of meeting the requirements of the United States after 1980. Since the late-1950's both military and civilian agencies have actively and independently pursued the idea of position determination and navigation using satellites. This resulted in the development of several systems with a multitude of specialized equipment responsive to particular mission requirements with varying degrees of accuracy and

Figure 32



The Measurement of the Geoid by the Satellite Altimeter

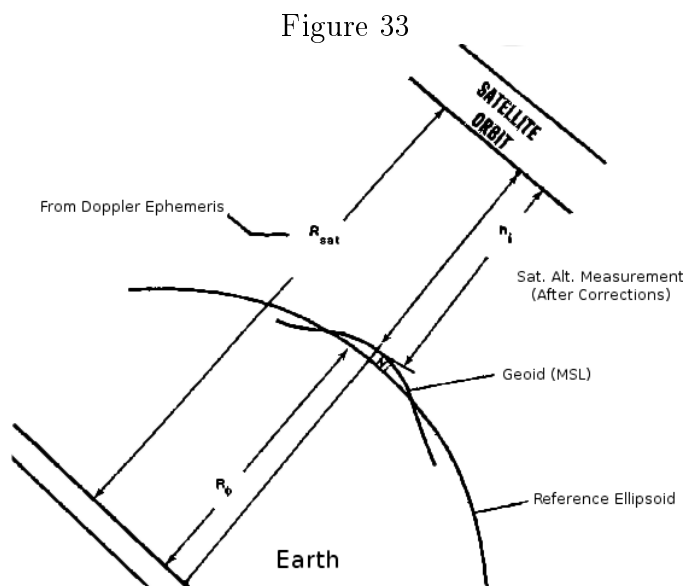
capabilities. In order to integrate the independent efforts of the military services, the Department of Defense issued a memorandum in 1973 naming the Air Force as the Executive Service for the initial development of a future Defense Navigation Satellite System (DNSS), designated the NAVSTAR Global Positioning System (GPS).

The GPS concept calls for a precise navigation system divided into three segments: space segment, control segment and user equipment segment. The space segment will consist of six orbital planes of satellites at inclinations of 55° in circular orbits at an altitude of 20,200 km. Each plane is to eventually contain three satellites. Each satellite will broadcast signals containing information as to its position. This broadcast will include an orbital ephemeris referenced to the DoD World Geodetic System. The control segment will be the ground stations necessary to track the satellites, monitor the system operation and periodically provide corrections to the navigation and time signals. The user segment will consist of the equipment necessary to convert the satellite signals into useful navigation information. By receiving signals from four satellites, the user, whether stationary or moving, can calculate his precise time, three-dimensional position and, if moving, his three-dimensional velocity. Position determination alone requires analysis of range information from three of the satellites in view. However, since the user's receiver clock will not be synchronized to the satellite clock, time of arrival measurements from four satellites are needed to update the user's clock.

When operational, GPS should satisfy the navigational accuracy requirements of many military-type missions on land, sea or in the air. Agencies also have many requirements for accurate geodetic positioning for which GPS will satisfy for years to come. These include establishing and densifying geodetic control, offshore positioning and the geodetic needs of national defense which brings in global requirements. GPS will also provide an excellent facility for determination of the position of other satellites and space vehicles while they are in lower earth orbits. This satellite-to-satellite tracking is discussed next.

Satellite-to-Satellite Tracking

A new technique for using artificial satellites for geodetic purposes is being studied and tested. This technique uses satellites to track other satellites. There are a number of variations which may be used for specific purposes such as gravity field investigations and orbit improvement. A high



altitude satellite may act as a relay from ground tracking stations to a low altitude satellite. In this way, low altitude satellites may be observed when they are not accessible to ground stations. In this type of tracking, a signal generated by a tracking station is received by the relay satellite and then retransmitted to a lower altitude satellite. This signal is then returned to the ground station by the same path. Two low altitude satellites can track one another observing mutual orbital variations caused by gravity field irregularities. Several high altitude satellites with accurately known orbits may be used to fix the position of a low altitude satellite. These examples present a few of the possibilities for the application of satellite-to-satellite tracking.

Satellite-to-satellite tracking data was first collected and analyzed in a high-low configuration between ATS-6 and GEOS-3. The data was studied to evaluate its potential for both orbit and gravitational model refinement. This experiment and others that followed proved this new technique to be an important tool for space geodesy.

Inertial Surveying

Inertial Navigation is the art and science of determining the position and velocity of a vehicle solely by means of sensing that vehicle's accelerations and performing the necessary integrations to determine the position and velocity on a real-time basis. The inertial system is composed of precise accelerometers to sense specific force acting on the vehicle and precise gyros to maintain orientation of the accelerometers in a chosen coordinate frame or to determine the orientation of the accelerometers with respect to that frame. Computation is performed by a small on-board computer and the position and velocity of the vehicle are displayed on a real-time basis. In the two decades that inertial navigation has been used, continued hardware developments have brought a state-of-the-art in which the inertially determined position of the vehicle is sufficiently accurate that inertial techniques can be applied to surveying.

At the heart of the inertial surveyor is the inertial measuring unit which contains three sensitive accelerometers and three precise gyros. The accelerometers are mounted as a mutually orthogonal triad on a platform which is torqued by the gyros to maintain orientation with the local vertical and local north, that is, the three axes are oriented north-east-down. The accelerometers measure the specific force on the vehicle which is the sum of the vehicle's own accelerations and the local gravity vector. The digitized output of the accelerometers are processed in real-time by a digital computer. They are integrated once to give velocity, and integrated again to give distance traveled along each sensitive axis. The system does not yield the latitude, longitude and elevation directly. To the computed distances, which are referenced to inertial space, there must be added the initial position and a conversion to latitude, longitude, and elevation accomplished. Although high quality accelerometers and gyros are used in the system, they are still subject to drift and bias. This will cause a misalignment of the platform and errors in the sensed accelerations, which results in small errors in computed velocities and positions. The currently available inertial surveying systems must stop or hover at frequent intervals. At these times, a Kalman filter process corrects for the difference between the indicated velocity and zero, and calculates normal gravity, elevation, and anomalous gravity, but only at these points where remaining errors in platform alignment are also corrected by the Kalman filter.

It was stated above that the accelerometers sense the sum of the vehicle's acceleration and the

local gravity vector and that the vehicle's accelerations are needed for integration into velocity and distance traveled. However, a model of the earth's gravity field is required to remove the accelerations due to gravity. In current systems, a very simplistic model is used in which only the downward gravity component resulting from an ellipsoidal earth is computed. Thus the system cannot correct for deflection of the vertical. Further, these deflections of the vertical result in erroneous platform alignments which may introduce errors as large as 40 cm in the computed positions.

Gravity Gradiometry

Gravity gradiometers have been suggested as a means of independently determining the components of the gravity vector on a real-time basis. A gravity gradient is simply the spatial derivative of the gravity vector. The gradient can be thought of as the rate of change of a component of the gravity vector as measured over a small distance. Hence, the gradient can be measured by determining the difference in gravity at two close but distinct points. This principle is embodied in several recent moving-base instruments. The gravity gradient at a point is a tensor, since it is the derivative of each component of the gravity vector taken in each sensitive axis. Thus, the value of any component of the gravity vector can be known all along the path of the vehicle if gravity gradiometers are included in the system and their outputs are integrated by the system computer. In theory, an accurate gravity model will be computed in real-time and a continuous map of normal gravity, elevation, and anomalous gravity will be available.

Chapter 8

The World Geodetic System

Because of the inability of the large geodetic systems such as the North American Datum (NAD), European Datum (ED), and Tokyo Datum (TD), to provide a basis for expression of intercontinental geodetic information, a unified world system became essential. The Department of Defense, in the late 1950's began to develop the needed world system to which geodetic datums could be referred and compatibility established between the coordinates of widely separated sites of interest. Efforts of the Army, Navy and Air Force were combined leading to the DoD World Geodetic System 1960 (WGS60). In accomplishing WGS60, a combination of available surface gravity data, astrogeodetic data and results from HIRAN and Canadian SHORAN surveys were used to define a best-fitting ellipsoid and an earth-centered orientation for each of the initially selected datums (Chapter 4). (The datums are relatively oriented with respect to different portions of the geoid by the astrogeodetic methods already described.) The sole contribution of satellite data to the development of WGS60 was a value for the ellipsoid flattening which was obtained from the nodal motion of a satellite.

Prior to WGS60, the Army and Air Force had each developed a world system by using different approaches to the gravimetric datum orientation method. To determine their gravimetric orientation parameters, the Air Force used the mean of the differences between the gravimetric and astrogeodetic deflections and geoid heights (undulations) at specifically selected stations in the areas of the major datums. The Army performed an adjustment to minimize the difference between astrogeodetic and gravimetric geoids. By matching the relative astro-geodetic geoids of the selected datums with an earth-centered gravimetric geoid, the selected datums were reduced to an earth-centered orientation. Since the Army and Air Force systems agreed remarkably well for the NAD, ED and TD areas, they were consolidated and became WGS60.

The Department of Defense World Geodetic System 1966

In January 1966, a World Geodetic System Committee composed of representatives from the Army, Navy and Air Force, was charged with the responsibility of developing an improved WGS needed to satisfy mapping, charting and geodetic requirements. Additional surface gravity observations, results from the extension of triangulation and trilateration networks, and large amounts of Doppler and optical satellite data had become available since the development of WGS60. Using the additional data and improved techniques, WGS66 was produced which served DoD needs for about five years after its implementation in 1967. The defining parameters of the WGS66 Ellipsoid

were the flattening $\frac{1}{298.25}$, determined from satellite data and the semi-major axis 6,378,145 meters, determined from a combination of Doppler satellite and astrogeodetic data. A worldwide $5^\circ \times 5^\circ$ mean free air gravity anomaly field provided the basic data for producing the WGS66 gravimetric geoid. Also, a geoid referenced to the WGS66 Ellipsoid was derived from available astrogeodetic data to provide a detailed representation of limited land areas.

The Department of Defense World Geodetic System 1972

After an extensive effort extending over a period of approximately three years, the Department of Defense World Geodetic System 1972 was completed. Selected satellite, surface gravity and astrogeodetic data available through 1972 from both DoD and non-DoD sources were used in a Unified WGS Solution (a large scale least squares adjustment). The results of the adjustment consisted of corrections to initial station coordinates and coefficients of the gravitational field.

The largest collection of data ever used for WGS purposes was assembled, processed and applied in the development of WGS72. Both optical and electronic satellite data were used. The electronic satellite data consisted, in part, of Doppler data provided by the U.S. Navy and cooperating non-DoD satellite tracking stations established in support of the Navy's Navigational Satellite System (NNSS). Doppler data was also available from the numerous sites established by GEOCEIVERS during 1971 and 1972. Doppler data was the primary data source for WGS72. Additional electronic satellite data was provided by the SECOR (Sequential Collation of Range) Equatorial Network completed by the U.S. Army in 1970. Optical satellite data from the Worldwide Geometric Satellite Triangulation Program was provided by the BC-4 camera system. Data from the Smithsonian Astrophysical Observatory was also used which included camera (Baker Nunn) and some laser ranging.

The surface gravity field used in the Unified WGS Solution consisted of a set of $410 \ 10^\circ \times 10^\circ$ equal area mean free air gravity anomalies determined solely from terrestrial data. This gravity field includes mean anomaly values compiled directly from observed gravity data wherever the latter was available in sufficient quantity. The value for areas of sparse or no observational data were developed from geophysically compatible gravity approximations using gravity-geophysical correlation techniques. Approximately 45 percent of the 410 mean free air gravity anomaly values were determined directly from observed gravity data.

The astrogeodetic data in its basic form consists of deflection of the vertical components referred to the various national geodetic datums. These deflection values were integrated into astrogeodetic geoid charts referred to these national datums. The geoid heights contributed to the Unified WGS Solution by providing additional and more detailed data for land areas. Conventional ground survey data was included in the solution to enforce a consistent adjustment of the coordinates of neighboring observation sites of the BC-4, SECOR, Doppler and Baker-Nunn systems. Also, eight geodimeter long line precise traverses were included for the purpose of controlling the scale of the solution.

FIGURE 38 DOPPLER SATELLITE GROUND STATIONS PROVIDING DATA FOR WGS72 DEVELOPMENT

Doppler Satellite Ground Stations Providing Data for WGS 72 Development.

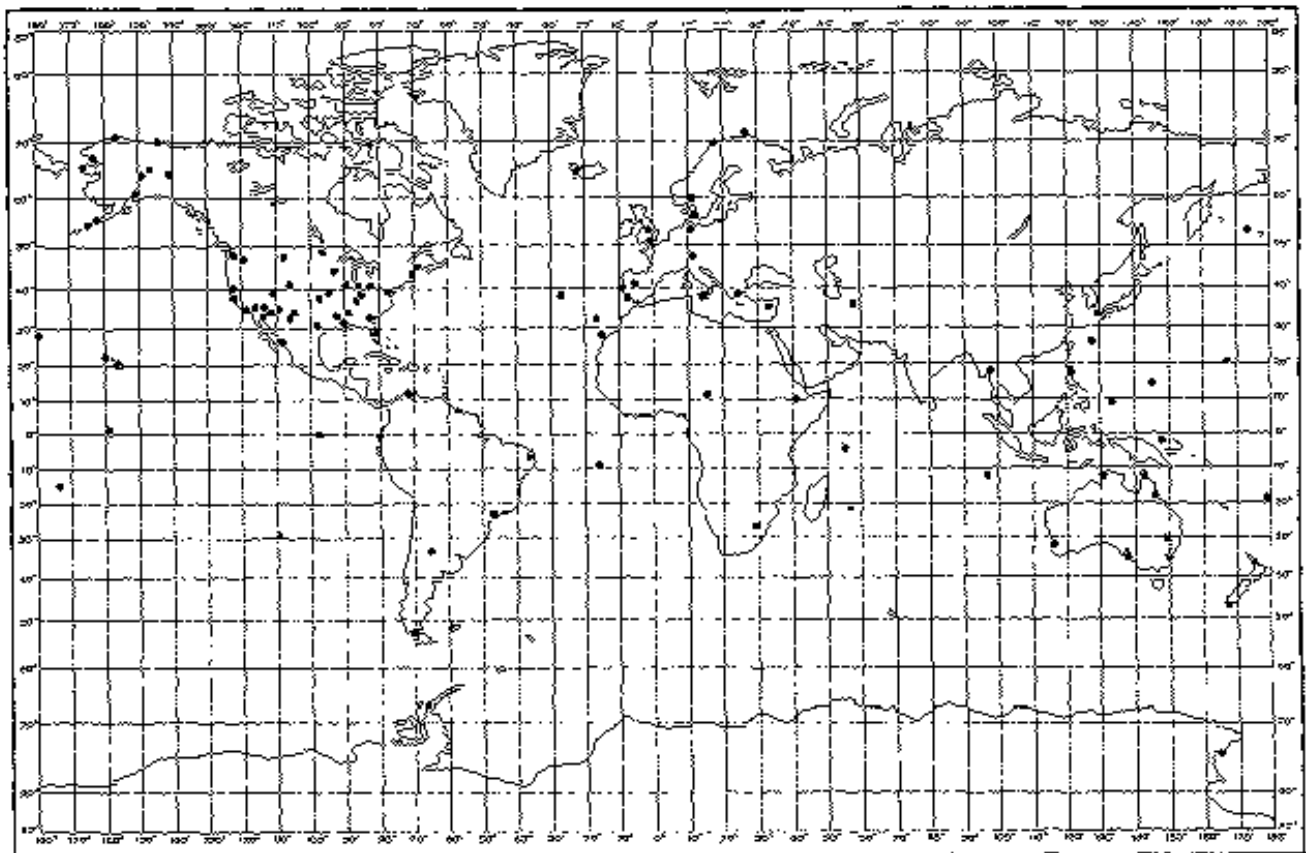
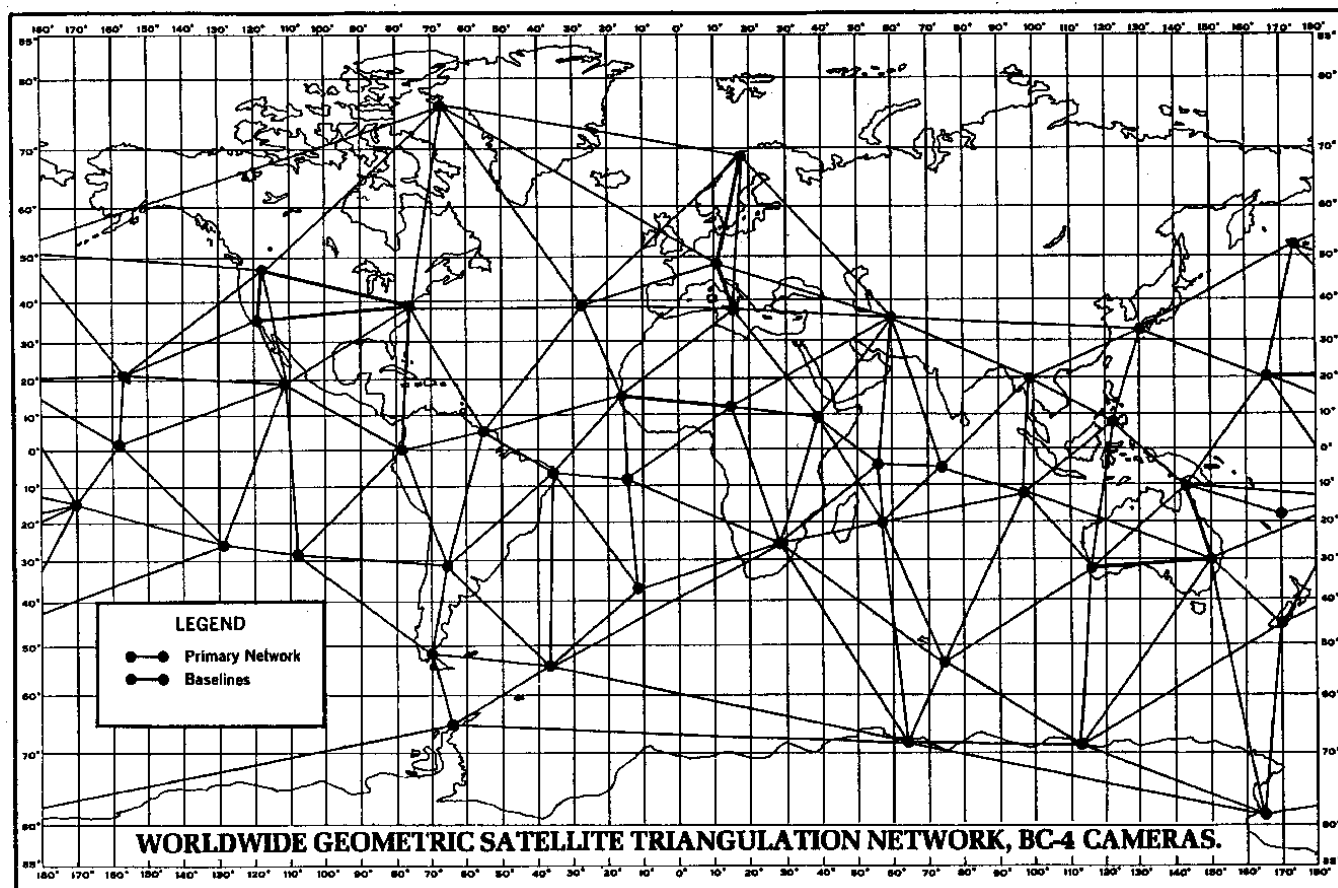


Figure 38

FIGURE 39 WORLDWIDE GEOMETRIC SATELLITE TRIANGULATION NETWORK, BC-4 CAMERAS



The Unified WGS Solution, as stated above, was a solution for geodetic positions and associated parameters of the gravitational field based on an optimum combination of available data. The WGS72 ellipsoid parameters, datum shifts and other associated constants were derived separately. For the unified solution, a normal equation matrix was formed based on each of the mentioned data sets. Then, the individual normal equation matrices were combined and the resultant matrix solved to obtain the positions and the parameters.

The value for the semi-major axis **a** of the WGS72 Ellipsoid is 6,378,135 meters. The adoption of an **a**-value 10 meters smaller than that for the WGS66 Ellipsoid was based on several calculations and indicators including a combination of satellite and surface gravity data for position and gravitational field determinations. Sets of satellite derived station coordinates and gravimetric deflection of the vertical and geoid height data were used to determine local to geocentric datum shifts, datum rotation parameters, a datum scale parameter and a value for the semi-major axis of the WGS Ellipsoid. Eight solutions were made with the various sets of input data, both from an investigative point of view and also because of the limited number of unknowns which could be solved for in any individual solution due to computer limitations. Selected Doppler satellite tracking and astro-geodetic datum orientation stations were included in the various solutions. Based on these results and other related studies accomplished by the Committee, an **a**-value of 6,378,135 meters and a flattening of $\frac{1}{298.26}$ were adopted.

In the development of local to WGS72 datum shifts, results from different geodetic disciplines were investigated, analyzed and compared. Those shifts adopted were based primarily on a large number of Doppler TRANET and GEOCEIVER station coordinates which were available worldwide. These coordinates had been determined using the Doppler point positioning method.

A New World Geodetic System

The need for a new world geodetic system is generally recognized by the geodetic community within and without the Department of Defense. WGS72 no longer provides sufficient data, information, geographic coverage, or product accuracy for all present and anticipated applications. The means for producing a new WGS are available in the form of improved data, increased data coverage, new data types and improved techniques. GRS80 parameters (Chapter 2), and presently available Doppler, laser and VLBI observations constitute significant new information which are being utilized. There are now over 1000 Doppler determined station positions available as compared to the near 100 station values used in WGS72. Also, an outstanding new source of data is now being made available from satellite radar altimetry (Chapter 7). Among the improved techniques now employed is an advanced least squares method called collocation which provides for a consistent solution from different types of measurements all relative to the earth's gravity field, i.e. geoid, gravity anomalies, deflections, dynamic Doppler, etc.

WGS (initially WGS72) is the reference system being used by the Global Positioning System. This has been updated to WGS84.

Chapter 9

Bibliography

1. Bomford, Brigadier G.; "Geodesy"; Oxford: Clarendon Press; London, England; 1952.
2. Brown, L.A.; "A Story of Maps"; Little, Brown and Company; Boston, Massachusetts; 1950.
3. Coast and Geodetic Survey; "Use of Near-Earth Satellite Orbits for Geodetic Information"; Technical Bulletin No. I 1; January 1960.
4. Coast and Geodetic Survey; "Use of Artificial Satellites for Navigation and Oceanographic Surveys"; Technical Bulletin No. 12; July 1960.
5. Guier, W.H. and G.C. Weiffenbach; "A Satellite Doppler Navigational System"; Proceedings of the IRE; Vol. 48, No. 4; April 1960; p. 507.
6. Guier, W.H. and G.C. Weiffenbach; "Space Navigation Handbook"; NAVPERS 92988.
7. Heiskanen, W.A. and F.A. Vening Meinesz; "The Earth and Its Gravity Field"; McGraw-Hill Book Company, Inc.; New York, New York; 1958.
8. Kaula, W.M.; "Celestial Geodesy"; NASA TN D-1 155; NASA; March 1962.
9. Kaula, W.M.; "SECOR Orbit Revealed"; Missiles and Rockets; 24 February 1964; p. 10.
10. Kershner, R.B. and R.R. Newton; "The Transit System"; The Journal of the Institute of Navigation; Vol. 15, No. 2; April 1962; pp. 129-149.
11. Coast and Geodetic Survey; "Control Leveling"; Special Publication No. 226; 1941.
12. Coast and Geodetic Survey; "Horizontal Control Data"; Special Publication No. 227; 1941.
13. Clark, D.; "Plane and Geodetic Surveying"; Constable and Company, Ltd.; London, England; 1957.
14. Hosmer, G.L.; "Geodesy"; John Wiley and Sons, Inc.; New York, New York; 1956.
15. Whitmore, G.D.; "Advanced Surveying and Mapping"; International Textbook Company; Scranton, Pennsylvania; 1952.
16. Nelson, E.H.; "The Smithsonian's Satellite-Tracking Program: Its History and Organization",. The Smithsonian Report for 1961; Smithsonian Institution Publication No. 4482; Washington, D.C.; 1962.
17. Seppelin, T.O.; "The Department of Defense World Geodetic System 1972"; Defense Mapping Agency (DMA) Aerospace Center; St. Louis AFS, Missouri; May 1974. [Prepared by the

World Geodetic System Committee, DMA; Washington, D.C.]

18. "DMA Report TR 75-001"; Department of Defense National Geodetic Satellite Program; DMA; Washington, D.C.; February 1975.

19. Smith, D.E., R. Kolenkiewicz, and P.J. Dunn; "Geodetic Studies by Laser Ranging to Satellites"; NASA Report No. X-553-71-361; Goddard Space Flight Center (GSFC); Greenbelt, Maryland; April 1971.

20. Smith, D.E. and R. Kolenkiewicz, et al; "The Applications of Laser Tracking to the Measurement of Intersite Distance"; Presented at the COSPAR Working Group 1]; Philadelphia, Pennsylvania; June 1976.

21. Wells, D.E.; "Semi Dynamical Doppler Satellite Positioning"; Bedford Institute of Oceanography; Dartmouth, Nova Scotia, Canada; August 1975.

22. Brown, D.C.; "Short Arc Optical Survey of the GEOS North American Tracking Network"; GSFC Document X-550-68-439; November 1968.

23. Gross, J.E.; "Preprocessing Electronic Satellite Observations; Reports of the Department of Geodetic Science", Report No. 100; The Ohio State University; Columbus, Ohio; March 1968.

24. Wells, D.E.; "Orbital Information from The U.S. Navy Navigation Satellite System"; Bedford Institute of Oceanography; Dartmouth, Nova Scotia, Canada; May 1971.

25. Mulholland, J.D., and C.O. Alley, et al; "Preliminary Results of Laser Ranging to a Reflector on the Lunar Surface"; Space Research XI; 1971.

26. Schwarz, C.R., and R.W. Smith; "Accuracies and Errors Associated with Miniaturized Doppler Receivers"; USA TOPOCOM (DMATC) Geodetic Memorandum No. 1670; Washington, D.C.; April 1972.

27. Gaposchkin, E.M.; "1973 Smithsonian Standard Earth (III)"; Smithsonian Institution Astrophysical Observatory; Cambridge, Massachusetts; November 1973.

28. Schmidt, Captain J.R.; "Computer Error Analysis of Tropospheric Effects for the NAVSTAR Global Positioning System"; Air Force Institute of Technology; Wright-Patterson AFB, Ohio; January 1975.

29. SAMSO Document; "Navigation Signal Structure of the NAVSTAR Global Positioning System"; SAMSO; Los Angeles, California.

30. Shapiro, I.I., and C.F. Martin; "Precision Geodesy Via Radio Interferometry: First Results"; Massachusetts Institute of Technology; Cambridge, Massachusetts; March 1971.

31. Shapiro, I.I., and A.E.E. Rogers; "Calibration of Very Long Baseline Interferometers for Geodetic Applications"; Northeast Radio Observatory Corporation; Cambridge, Massachusetts; February 1973.

32. "IAG Special Publication No. 3; Geodetic Reference System 1967"; Central Bureau IAG; Paris, France; 1971.

33. "National Ocean Survey Publication; Classification, Standards of Accuracy, and General Specifications of Geodetic Control Surveys"; National Oceanic and Atmospheric Administration; Rockville, Maryland; February 1974.

34. NASA Publication; "NASA Directory of Observation Station Locations, Vol. 1"; Computer Science Corporation; Falls Church, Virginia; November 1973.
35. Meade, B.K.; "High-Precision Transcontinental Traverse Surveys in the United States"; National Oceanic and Atmospheric Administration; Rockville, Maryland; 1969.
36. Smith, P.A.; "North American Datum"; Report of the NAS/NAE Committee on the North American Datum; National Academy of Sciences; Washington, D.C.; December 1969.
37. "DMA Technical Manual T-3-52320; Satellite Records Manual Doppler Geodetic Point Positioning, Data Documentation and Applications"; DMA; Washington, D.C.; November 1976.
38. Wilcox, L.E.; "An Analysis of Gravity Prediction Methods for Continental Areas; DMAAC Reference Publication No. 74-001"; DMA Aerospace Center; St. Louis AFS, Missouri; August 1974.
39. Brown, D.C.; "Doppler Positioning by the Short Arc Method"; DBA Systems, Inc.; Melbourne, Florida; October 1976.
40. Ewing, C.E. and M.M. Mitchell; "Introduction to Geodesy"; American Elsevier Publishing Company, Inc.; New York, New York; 1970.
41. Heiskanen, W.A. and H. Moritz; "Physical Geodesy"; W.H. Freeman and Company; San Francisco, California; 1967.
42. Mueller, I.I. and J.D. Rockie; "Gravimetric and Celestial Geodesy"; Frederick Ungar Publishing Co.; New York, New York; 1966.
43. Kaula, W. M.; "Theory of Satellite Geodesy"; Blaisdell Publishing Company; Waltham, Massachusetts; 1966.
44. Phillips, J.O.; "Specifications to Support Classification, Standards of Accuracy, and General Specifications of Geodetic Control Surveys"; Federal Geodetic Control Committee; National Oceanic and Atmospheric Administration; Rockville, Maryland; 1980.

Appendix A

Spherical Harmonics

A. HARMONIC EXPRESSIONS

Trigonometric functions, sines and cosines, are harmonic in that they repeat themselves after a certain interval. Harmonic Expressions, formed by combining sines, cosines and arbitrary constants, are mathematical devices for curve-fitting and interpolation in either two or three dimensions. An example of a harmonic Expression for a two-dimensional curve is

$$Y = A_0 + A_1\cos(\Theta) + A_2\sin(\Theta) + A_3\cos(2\Theta) + A_4\sin(2\Theta) + \dots \quad (\text{A.1})$$

where A_0, A_1, A_2, \dots , are arbitrary constants.

The nature of every harmonic expression is that the constant term (A_0 in the equation above) equals the mean or average value of all points on the curve. The succeeding terms represent harmonic deviations from that mean value. Each harmonic term has two parts: the trigonometric part which is harmonic and controls the frequency of the oscillations, and a constant multiplier which controls the amplitude. For the second term in equation (1)

$$A_1\cos(\Theta)$$

$\cos(\Theta)$ is the trigonometric part and A_1 is the constant multiplier.

Example. Consider the harmonic expression

$$Y = 3 + 2\cos(\Theta) + 2\sin(\Theta) + 4\cos(2\Theta) - \sin(8\Theta)$$

which describes the curve shown in Figure A-1. Its first term, which is its constant term, is 3. Thus 3 is the mean value of the curve. The second term, $y = 2\cos(\Theta)$, is a harmonic of period 360° , because after this period it repeats itself. Its amplitude is 2, the largest value for y this term can attain. The third term, $y = 2\sin(\Theta)$, traces the same curve as $y = 2\cos(\Theta)$ except that it is 90° out

of phase, or shifted 90° . The term $y = 4\cos(2\Theta)$ has an amplitude of 4 and a frequency of 2. The frequency of 2 means that the curve for this term repeats itself twice every 360° . The last term, $y = -\sin(8\Theta)$, has an amplitude of -1 and a frequency of 8. The negative constant indicates that the curve for this term will start on the negative side of $y = 0$, rather than on the positive side. Graphs of the harmonic terms are shown in Figure A-2.

B. SPHERICAL HARMONICS

In the geosciences, a commonly used form of harmonics are the surface spherical harmonics. These functions are much used in geodesy because they are relatively simple to use and because the earth's shape can be approximated by a sphere. The coordinates normally associated with spherical harmonics are the spherical coordinates: radius r , co-latitude or polar distance (Θ), and longitude (λ).

1. Legendre Functions

The basic spherical harmonic form is the Legendre polynomial, $P_n(\cos(\Theta))$, given by the Formula:

$$P_n(\cos(\Theta)) = \frac{1}{2^n!} \frac{d^n(\cos^2(\Theta) - 1)^n}{d(\cos(\Theta))^n} \quad (\text{A.2})$$

where n , a nonnegative integer, is the degree of the polynomial. The Legendre polynomials for degrees $n = 0$ through $n = 5$ are:

$$P_0(\cos(\Theta)) = 1$$

$$P_1(\cos(\Theta)) = \cos(\Theta)$$

$$P_2(\cos(\Theta)) = \frac{1}{2} (3 \cos^2(\Theta) - 1)$$

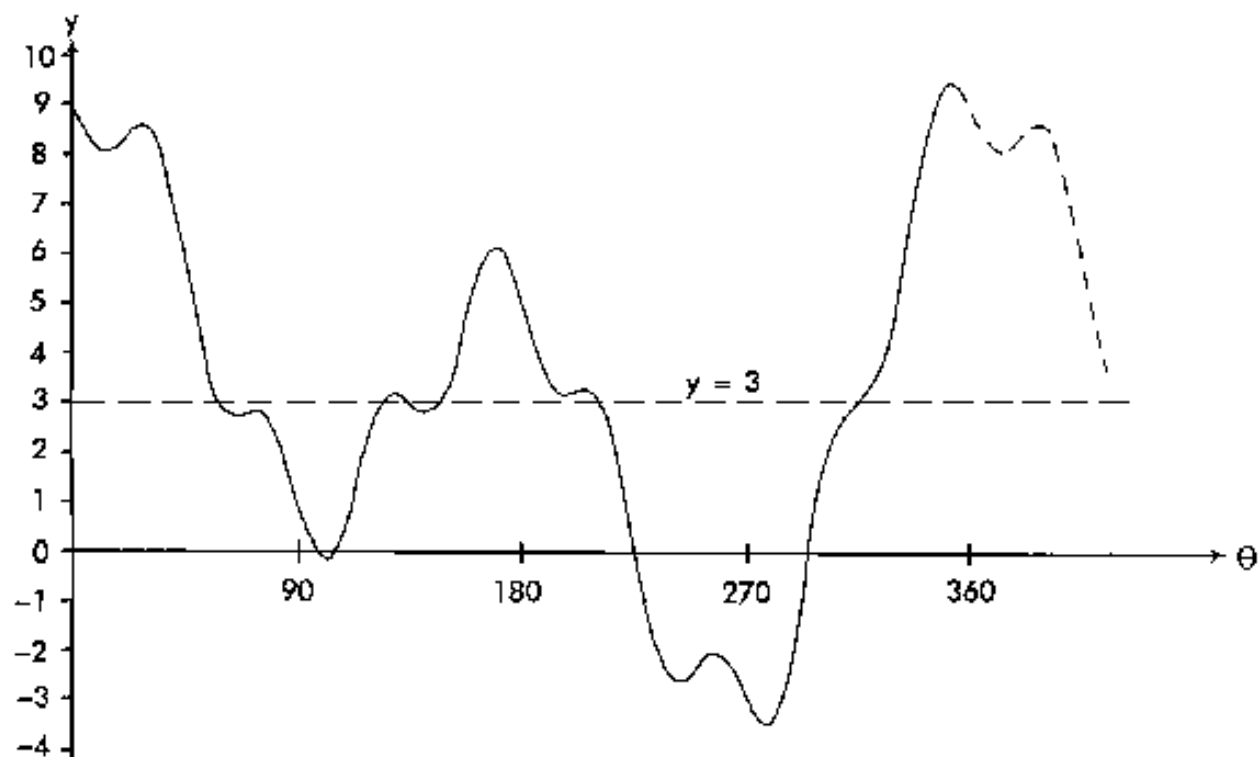
$$P_3(\cos(\Theta)) = \frac{1}{2} (5 \cos^3(\Theta) - 3 \cos(\Theta))$$

$$P_4(\cos(\Theta)) = \frac{1}{8} (35 \cos^4(\Theta) - 30 \cos^2(\Theta) + 3)$$

$$P_5(\cos(\Theta)) = \frac{1}{8} (63 \cos^5(\Theta) - 70 \cos^3(\Theta) + 15 \cos(\Theta))$$

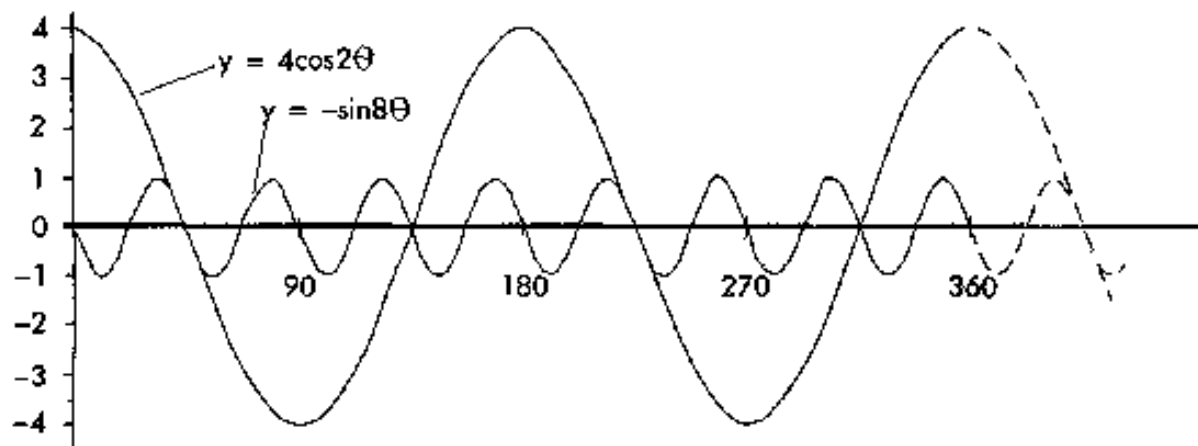
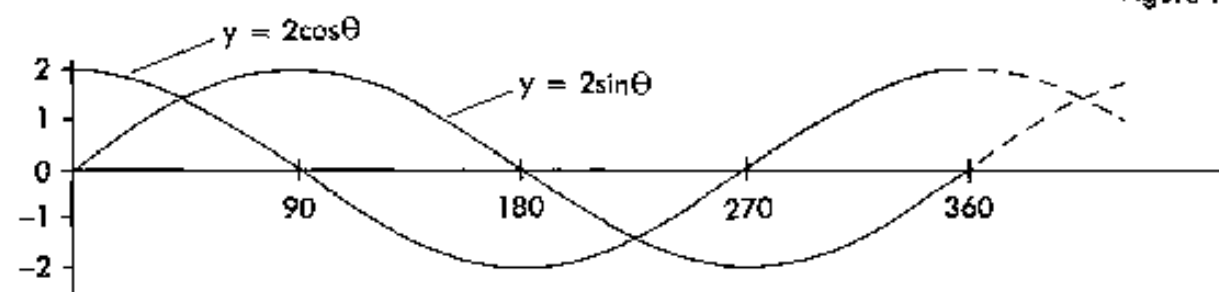
Graphs of these polynomials up to $n = 4$ are shown in Figure A-3.

Figure A1 & A2



HARMONIC EXPRESSION $y = 3 + 2\cos\theta + 2\sin\theta + 4\cos 2\theta - \sin 8\theta$

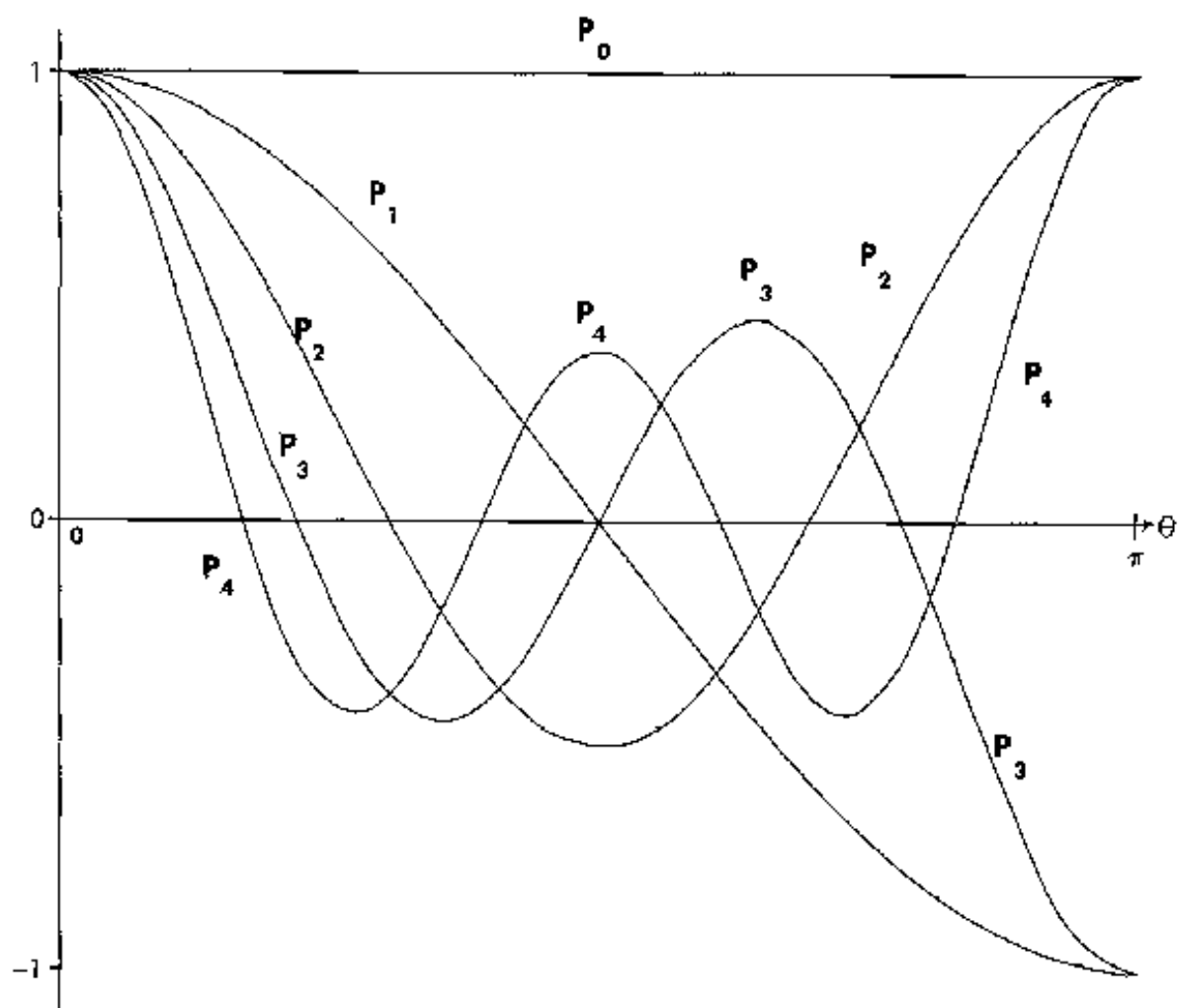
Figure A-1



HARMONIC TERMS: $2\cos\theta, 2\sin\theta, 4\cos\theta, -\sin 8\theta$

Figure A-2

Figure A3

**LEGENDRE POLYNOMIALS, P_0 TO P_4**

Another form of spherical harmonics are the associated Legendre functions, $P_{nm} \cos(\Theta)$, where m , a nonnegative integer smaller than or equal to n , is the order of the function. When $m = 0$, this function becomes $P_{n0} \cos(\Theta)$, or simply the Legendre polynomial $P_n \cos(\Theta)$.

The surface spherical harmonics are Legendre functions multiplied by $\cos m(\lambda)$ or $\sin m(\lambda)$, where λ is the longitude,

degree 0: $P_0(\cos(\Theta))$

degree 1: $P_1(\cos(\Theta))$

$$P_{11}(\cos(\Theta))\cos(\lambda), P_{11}(\cos(\Theta))\sin(\lambda)$$

degree 2: $P_2(\cos(\Theta))$

$$P_{21}(\cos(\Theta))\cos(\lambda), P_{21}(\cos(\Theta))\sin(\lambda)$$

$$P_{22}(\cos(\Theta))\cos(2\lambda), P_{22}(\cos(\Theta))\sin(2\lambda)$$

and so on.

These resultants are tesseral harmonics, which divide the sphere into blocks of alternate positive and negative values, called tesserals. Figure A-4 shows a typical tesseral harmonic. The shaded areas represent tesserae where the harmonic is positive, and the non-shaded areas represent negative tesserae.

A special case arises when $m = 0$, in which case the $\sin m(\lambda)$ harmonic terms disappear because $\sin 0(\lambda) = 0$. When $m = 0$, the tesseral harmonic divides the sphere into zones of alternate positive and negative values, and is therefore called a zonal harmonic. An additional special case arises when $m = n$. Then the tesseral harmonic divides the sphere into sectors of alternate positive and negative values. Hence they are called sectorial harmonics. Examples of zonal and sectorial harmonics are shown in Figure A-4.

2. Surface Spherical Harmonics

The general expression for the surface spherical harmonic of an arbitrary function $f((\Theta),(\lambda))$ is

$$f(\Theta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^n [A_{nm} P_{nm}(\cos\Theta) \cos(m\lambda) + B_{nm} P_{nm}(\cos\Theta) \sin(m\lambda)] \quad (\text{A.3})$$

where A_{nm} and B_{nm} are arbitrary constants. For simplicity and easier handling, the harmonics in Equation (3), commonly referred to as conventional harmonics, can be replaced by other functions called fully normalized harmonics. The surface spherical harmonic then becomes

$$f(\Theta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^n [\bar{A}_{nm} \bar{P}_{nm}(\cos\Theta) \cos(m\lambda) + \bar{B}_{nm} \bar{P}_{nm}(\cos\Theta) \sin(m\lambda)] \quad (\text{A.4})$$

where the bar indicates that the harmonic term is fully normalized. The conventional and fully normalized terms are related through the following expressions:

$$\bar{A}_{n0} = \frac{1}{2n+1} A_{n0}$$

$$\bar{A}_{nm} \quad (\text{A.5})$$

$$\bar{P}_{nm} \quad (\text{A.6})$$

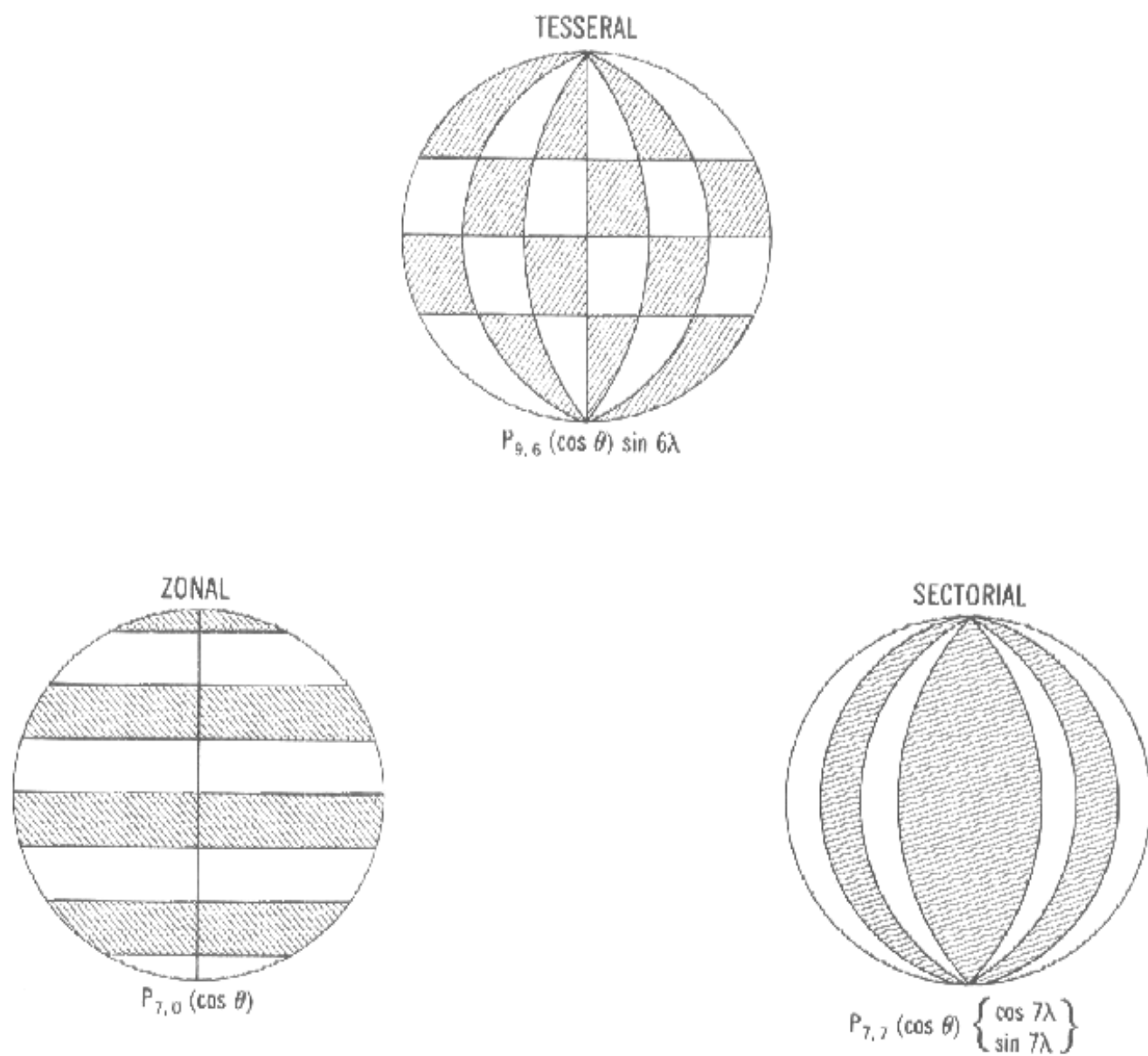
In many applications, geopotential coefficients, symbolized by C and S, are used in harmonic expressions rather than the A and B coefficients. Like the A and B coefficients, they can be used in either conventional or fully normalized form. The relationship between the two types of coefficients is.

$$C_{nm} = \frac{A_{nm}}{GMa^n} \quad (\text{A.7})$$

$$S_{nm} = \frac{B_{nm}}{GMa^n}$$

where a is the equatorial radius and GM is the gravitational constant. Fully normalized geopotential coefficients, C and S , can be obtained from the conventional coefficients, \bar{C} and \bar{S} , by means of an expression similar to Equation (5).

Figure A4



**EXAMPLES of THE DIFFERENT KINDS
of SPHERICAL HARMONICS**

C. GRAVITATIONAL POTENTIAL OF THE EARTH

The expression for the earth's gravitational potential, that is, the potential of the attraction of the earth, is:

$$U = \frac{GM}{r} \left[1 + \sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{a}{r}\right)^n (\bar{C}_{nm} \cos(m\lambda) + \bar{S}_{nm} \sin(m\lambda)) \bar{P}_{nm} \cos(\Theta) \right] \quad (\text{A.8})$$

Where

U = gravitational potential function

GM = Earth's gravitational constant

= product of the universal gravitational constant and the earth's mass

r = radius vector from the earth's center of gravity

a = semimajor axis or equatorial radius

n, m = degree and order, respectively

Θ = polar distance of colatitude

λ = longitude

$\bar{C}_{nm}, \bar{S}_{nm}$ = normalized harmonic coefficients

$\bar{P}_{nm} \cos(\Theta)$ = normalized associated Legendre function.

The ability to express the gravitational potential by Equation (8) lies in determining proper values for the harmonic coefficients, \bar{C} and \bar{S} . These coefficients are a function of the mass distribution of the earth, which unfortunately is not known. Thus, the coefficients have to be computed in some other way, such as, from artificial satellite perturbations or gravity observations on the surface of the earth.

In the expression for the gravitational potential, the constant term, $U = \frac{GM}{r}$, which is the $n = 0$ term, equals the mean value. The mean value of the earth's gravitational potential is the value which would occur if the earth were a sphere. As in the two-dimensional case, each term after the mean value represents a harmonic deviation from that mean value. As was previously stated, the magnitude of the coefficients, \bar{C} and \bar{S} , is related to the mass distribution of the earth. Thus, where the potential is positive with respect to the mean value, there is a mass excess. Similarly, there is a mass deficiency where the potential is negative.

D. ANALYSIS OF LOWER DEGREE HARMONIC TERMS

The following discussion will analyze in detail some of the lower degree harmonic terms. In order to simplify writing, conventional harmonics will be used in this portion of the discussion.

Equation (8), expressed in terms of conventional harmonics, is

$$U = \frac{GM}{r} \left[1 + \sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{a}{r} \right)^n (C_{nm} \cos(m\lambda) + S_{nm} \sin(m\lambda)) P_{nm} \cos\Theta \right]$$

where

If this equation is expanded, the harmonic terms for degrees 2 and 1 are:

degree 1:

$$m = 0 \quad C_{10} P_{10}(\cos\Theta)$$

$$m = 1 \quad C_{11} P_{11}(\cos\Theta \cos\lambda), S_{11} P_{11}(\cos\Theta \sin\lambda)$$

degree 2:

$$m = 0 \quad C_{20} P_{20}(\cos\Theta)$$

$$m = 1 \quad C_{21} P_{21}(\cos\Theta \cos\lambda), S_{21} P_{21}(\cos\Theta \sin\lambda)$$

$$m = 2 \quad C_{22} P_{22}(\cos\Theta \cos\lambda), S_{22} P_{22}(\cos\Theta \sin\lambda)$$

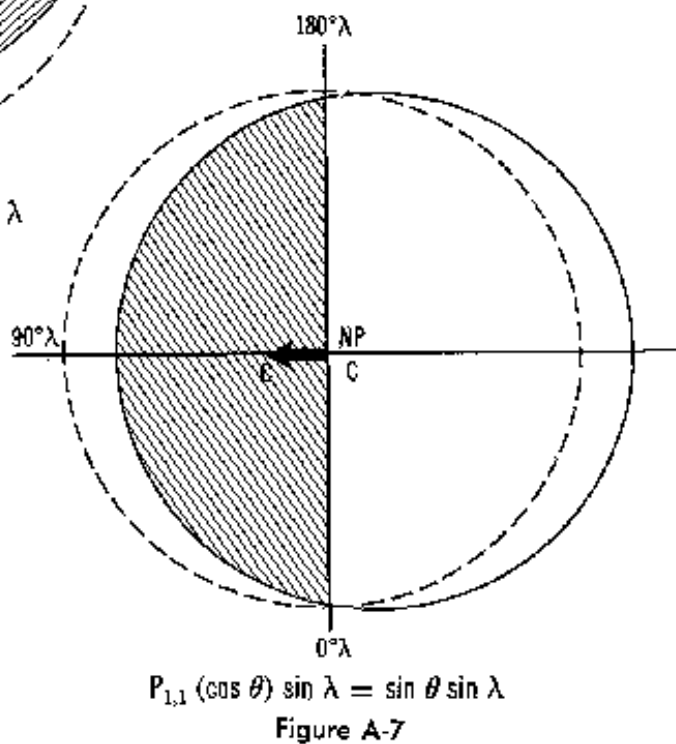
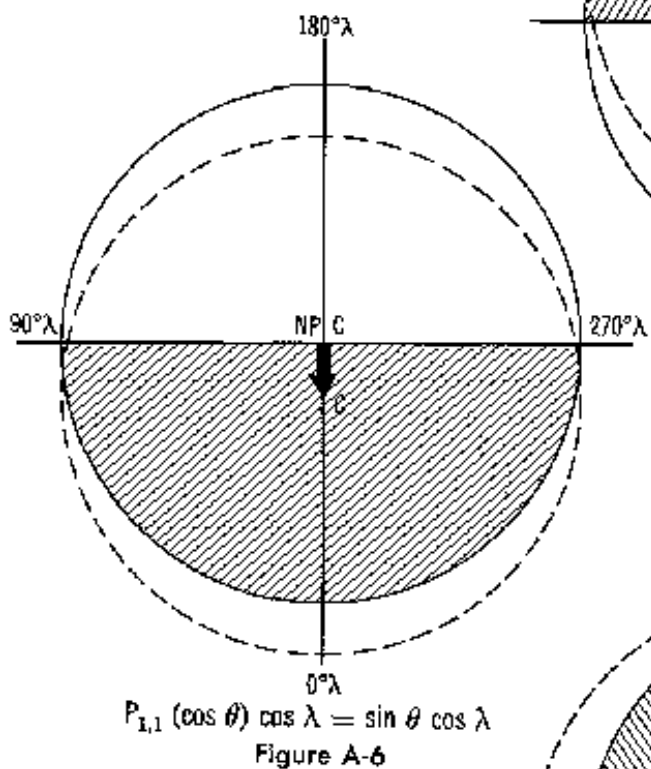
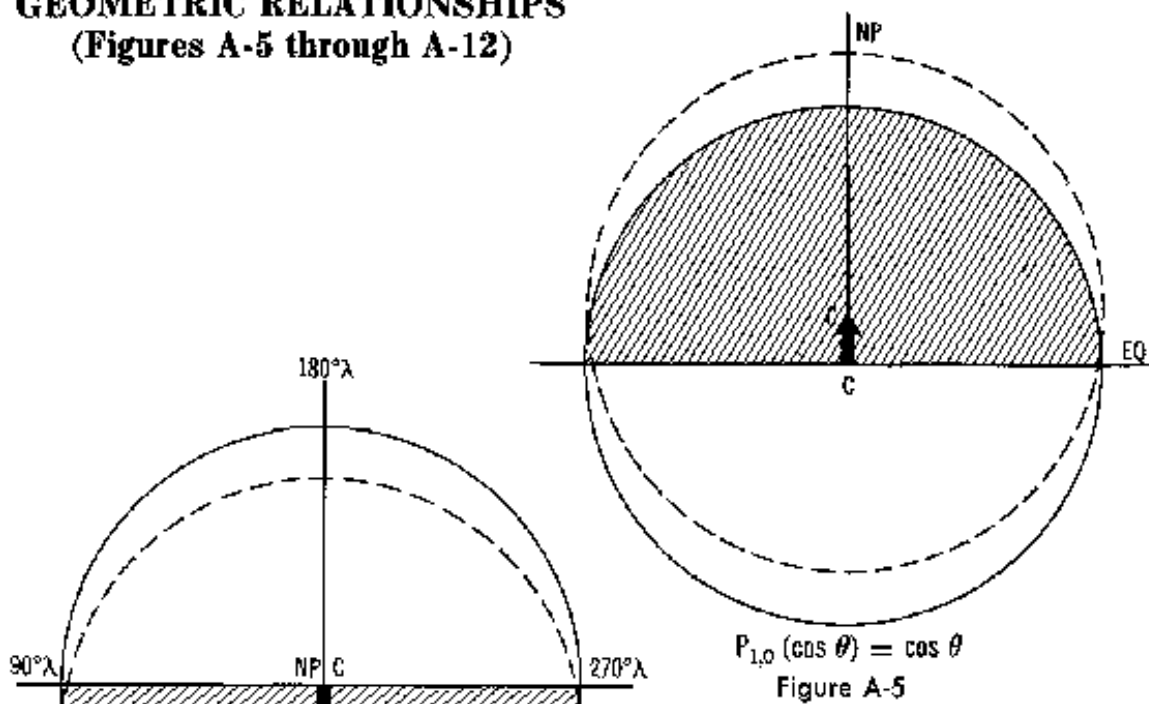
1. Degree $n = 1$

Figure A-5 shows a geometric representation of the harmonic term $P_{10} \cos(\Theta)$. The solid circle represents the mean value and the dashed curve is the value of the potential when only the mean value term and the term under consideration are combined. Thus, the dashed curve shows the effect of an individual term. The hachured part of the figure indicates the area of the sphere where the harmonic is positive; and the non-hachured part indicates where the harmonic is negative.

Since $\cos\Theta$ is positive from $\Theta = 0^\circ$ to $\Theta = 90^\circ$ and negative from $\Theta = 0^\circ$ to $\Theta = -90^\circ$, it follows that the harmonic $P_{10} \cos(\Theta) = \cos(\Theta)$ is positive in the northern hemisphere and negative in the southern. As is shown in Figure A-5, this corresponds to there being a mass excess in the northern hemisphere and a mass deficiency in the southern hemisphere. If this were true of the earth, the center of mass would be north of the center of coordinates. Since it is always specified that the center of coordinates and the center of mass coincide, this harmonic is not admissible in the potential expression and therefore must equal zero. In order to make this term zero, the coefficient C_{10} is set equal to zero, so that the entire term disappears.

Figure A-5 through A-12

GEOMETRIC RELATIONSHIPS
(Figures A-5 through A-12)



Figures A-5, A-6, A-7

The $P_{11} \cos\Theta$ terms are shown in Figures A-6 and A-7. They represent hemispheres (eastern and western) of positive and negative values. Existence of the terms indicate a shift of the center of mass away from the center of the coordinates in the positive hemisphere. Each term represents different hemispheres of positive and negative values, shifted by 90° , since $\cos(x)$ and $\sin(x)$ are 90° out of phase. The same logic which applied to the harmonic $P_{10} \cos \Theta$ applies here and these harmonics are also inadmissible. Thus, their coefficients, C_{11} and S_{11} , are set equal to zero, so that the terms disappear.

2. DEGREE $N = 2$

The zonal harmonic $P_{20} \cos\Theta$ is shown in Figure A-8. In the figure, the constant C_{20} is assumed to be negative, thus representing a mass excess in the equatorial region and a mass deficiency in the polar regions. Such a mass distribution indicates that the earth must be flattened somewhat. Actually, the magnitude of the C_{20} coefficient is 100 times greater than that of any of the other coefficients. This indicates the extent to which the earth is flattened and the importance of this harmonic.

The next harmonic term in order is the $P_{21} \cos(\Theta) \cos(\lambda)$ term shown in Figure A-9. As indicated by the figure, it represents opposite northern and southern quadrants of mass excesses and deficiencies. This would make the axis of inertia revolve around the axis of rotation of the earth, thus causing the earth to wobble on its axis. Since there is really some wobbling, which is commonly referred to as precession, the coefficient C_{21} cannot equal zero. However, the precession is so small the C_{21} is also very small. In many cases it is approximated by zero and the term disappears. There is also a $\sin(\lambda)$ term which has similar geometry, except that the quadrants are shifted by 90° in longitude. The same logic applies.

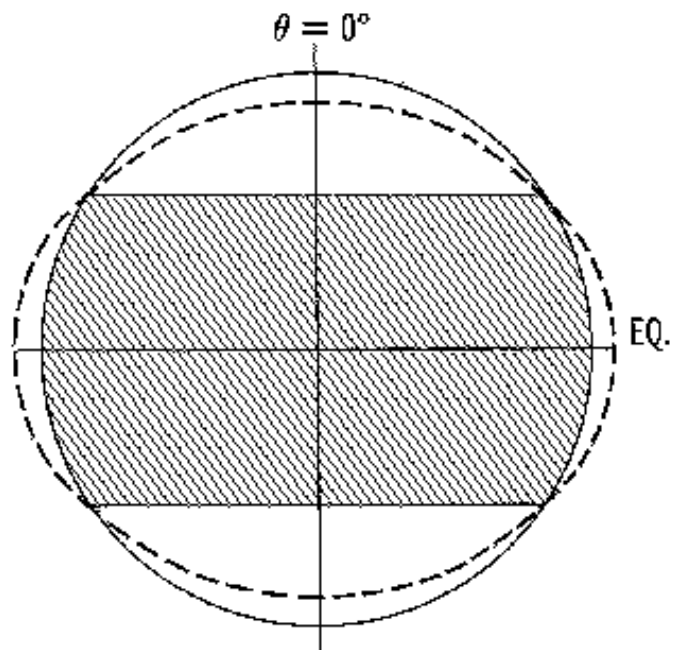
The next two terms are the sectoral harmonics $P_{22} \cos(\Theta)\cos(2\lambda)$ and $P_{22} \cos(\Theta)\sin(2\lambda)$, shown in Figure A-10. The figure shows sector quadrants which would indicate ellipticity of the equator. Each term represents ellipticity in a different direction. Thus, the two terms must be combined to obtain the overall ellipticity of the equator. The direction of the semi-major axis of the equatorial ellipse is at the point where their sum is a maximum. The direction of the semi-minor axis is 90° from the semi-major axis. As shown in Figure A-10, the sum is a maximum where the two curves meet. This is only true if their coefficients, C_{22} and S_{22} , are equal. The value for the ellipticity of the equator can be determined from the values of the coefficients C_{22} and S_{22} .

3. DEGREES $N = 3$ AND $N = 4$

The presence of the zonal harmonic $P_{30} \cos(\Theta)$ indicates that the earth is pear-shaped, as shown in Figure A-11. For this reason, this term has been referred to as the pear-shaped harmonic. Since values have been computed for its coefficient C_{30} , the earth is considered to be somewhat pear-shaped.

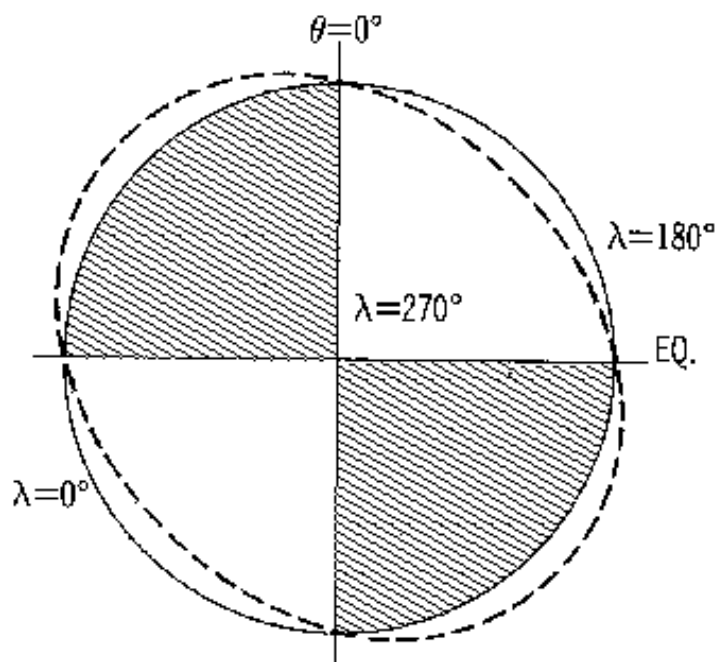
The last harmonic term that will be mentioned here is the zonal harmonic $P_{40} \cos \Theta$, shown in Figure A-12. Although it has no recognizable geometric relationship, it is important because it is a term of the normal gravity formula and permits refinement of the potential on an ellipsoid.

Figure A-8 through A-9



$$P_{2,0}(\cos \theta) = \frac{1}{2} (3 \cos^2 \theta - 1)$$

Figure A-8



$$P_{2,1}(\cos \theta) \cos \lambda = 3 \sin \theta \cos \theta \cos \lambda$$

Figure A-9

Figure A-10

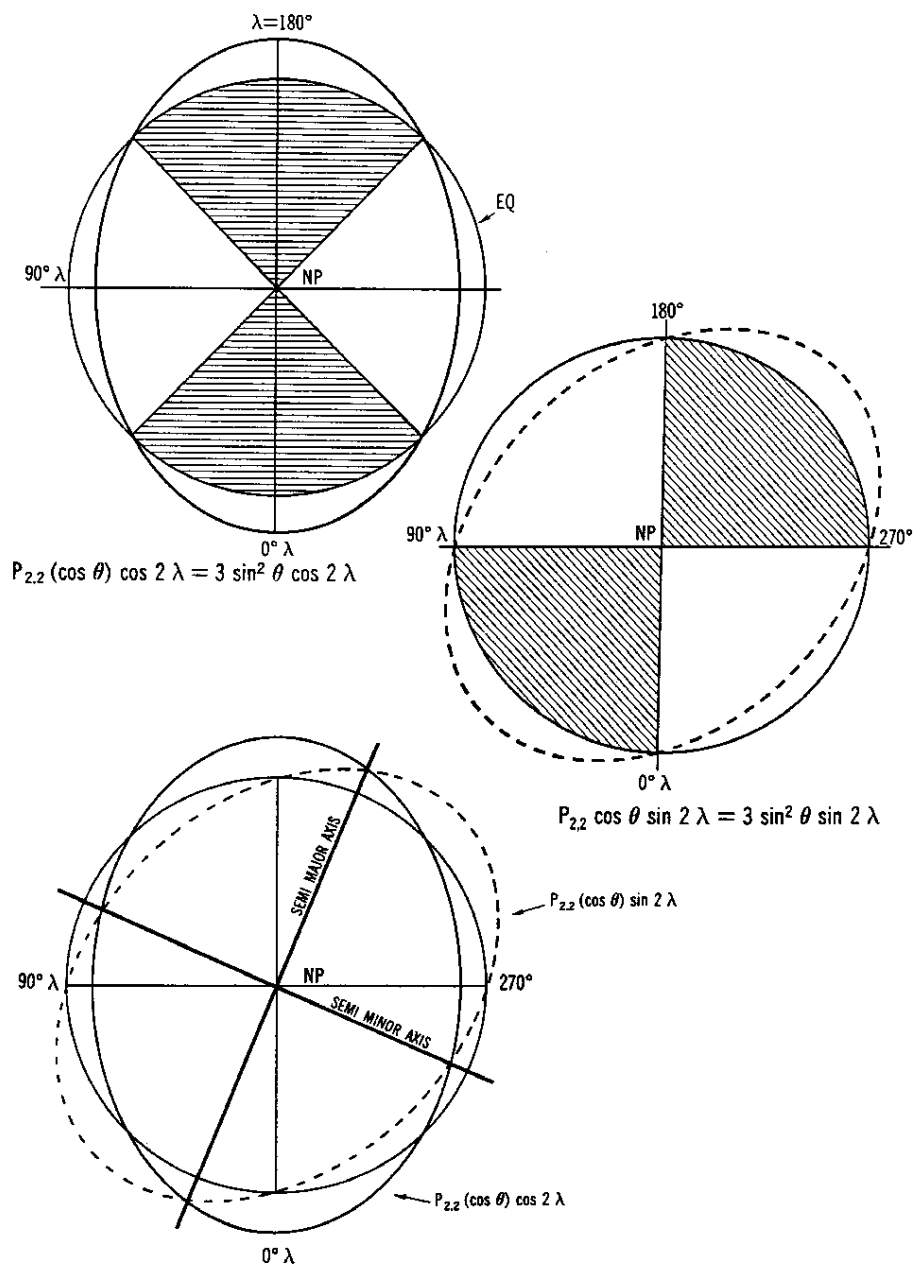


Figure A-11 through A-12

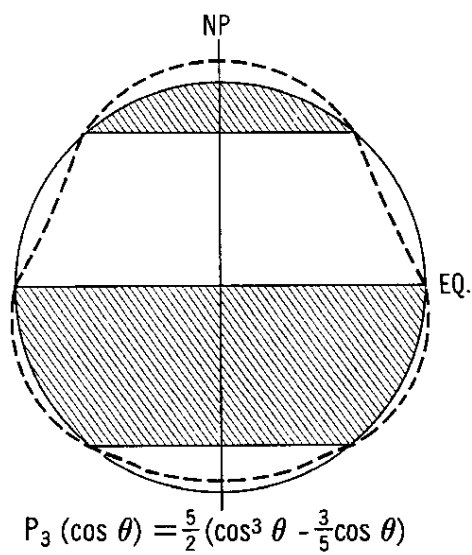


Figure A-11

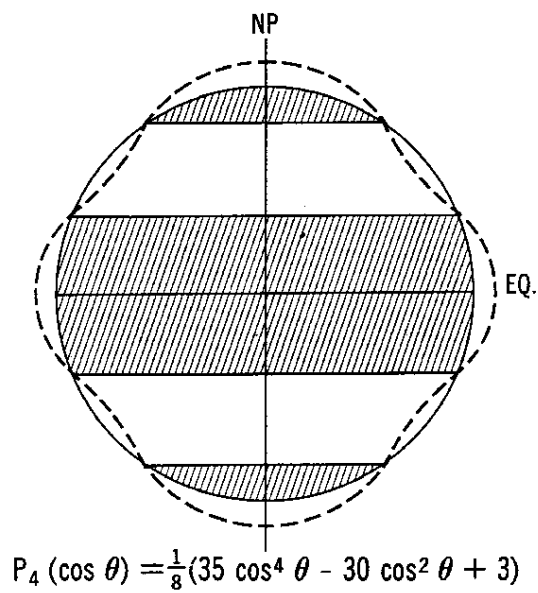


Figure A-12

E. APPLICATIONS OF SPHERICAL HARMONICS

There are an unlimited number of functions which can be expressed in terms of spherical harmonics. One of these functions is the normal gravitational potential of the reference ellipsoid.

The normal gravitational potential V , which is the gravity potential of the reference ellipsoid minus the potential of centrifugal force, can be developed into a series of zonal spherical harmonics.

$$V = \frac{GM}{r} \left\langle 1 - \sum_{n=1}^{\infty} J_{2n} \left(\frac{a}{r} \right)^{2n} \cdot P_{2n} \cos(\Theta) \right\rangle \quad (\text{A.9})$$

where

GM = earth's gravitational constant

r = radius vector

a = equatorial radius

J_{2n} = constant coefficients

For this application, the J coefficients are usually used rather than the C coefficients. There is a constant relationship between the different coefficients.

In some applications involving the reference ellipsoid, equation 9 is truncated at $n = 2$, giving

$$V = \frac{GM}{r} \left[1 - J_2 \left(\frac{a}{r} \right)^2 P_2 \cos \Theta - J_4 \left(\frac{a}{r} \right)^4 P_4 \cos \Theta \right]$$

or

$$V = \frac{GM}{r} \left[1 - J_2 \left(\frac{a}{r} \right)^2 (3 \cos^2 \Theta - 1) - \frac{1}{8} J_4 \left(\frac{a}{r} \right)^4 (35 \cos^4 \Theta - 30 \cos^2 \Theta + 3) \right]$$

Mathematical expressions for the J_2 and J_4 coefficients are

$$J_2 = \frac{2}{3}f - \frac{1}{3}m - \frac{1}{3}f^2 + \frac{2}{21}fm \quad (\text{A.10})$$

$$J_4 = -\frac{4}{5}f^2 + \frac{4}{7}fm$$

where

f = flattening

$m = \frac{\omega^2 a^2 b}{GM}$ = ratio of the centrifugal force to gravity at the equator.

ω = angular velocity of the earth's rotation.

b = semi-major axis.

Equation 10 shows the relationship between the flattening and the J_2 coefficient. An example of a geodetic application of artificial earth satellites is to determine J_2 from satellite perturbations, and to use that value in equation 10 to determine the flattening.